

**Methods and Compositions for Sensitive and Rapid, Functional Identification of  
Genomic Polynucleotides And Use For Cellular Assays in Drug Discovery**

5

**Cross Reference to Related Applications**

Under 35 USC § 120, this application claims the benefit of prior U.S. Application  
Serial No. 09/047,862, filed March 25, 1998, which is a continuation-in-part of U.S. Patent  
Application Serial No. 09/021,974, filed February 11, 1998, which is a continuation-in-part  
of PCT/US97/17395, filed September 26, 1997, which is a continuation-in-part of  
08/719,697, filed September 26, 1996, the contents of which are incorporated by reference  
in their entirety herein.

10

**Technical Field**

The present invention generally relates to methods and composition sfor the  
identification of useful and functional portions of the genome and compounds for  
modulating such portions of the genome. The present invention particularly relates to the  
use of viral vectors, such as adeno-associated viruses (AAV) and retroviruses to identify  
useful and functional portions of the genome, such as genes and promoters.

15

20

**Background**

The identification and isolation of useful portions of the genome requires  
extensive expenditure of time and financial resources. Currently, many genome projects  
use various strategies to reduce cloning and sequencing times. While genome projects

25

rapidly expand the database of genetic material. such projects often lack the ability to integrate the information with the biology of the cell or organism from which the genes were isolated. In some instances, coding regions of newly isolated genes reveal sequence homology with other genes of known function. This type of analysis can, at best, provide clues to the possible relationships between different genes and proteins. Genomic projects in general, however, suffer from the inability to rapidly and directly isolate, and identify specific, yet unknown, genes associated with particular a biological process or processes.

The evaluation of the function of genes identified from genomic sequencing projects requires cloning the discovered gene into an expression system suitable for functional screening. Transferring the discovered gene into a functional screening system requires additional expenditure of time and resources without a guarantee that the correct screening system was chosen. Since the function of the discovered gene is often unknown or only surmised by inference to structurally related genes, the chosen screening system may not have any relationship to the biological function of the gene. For example a gene may encode a protein that is structurally homologous to the beta-adrenergic receptor and have a dissimilar function. Further, if negative results are obtained in the screen, it can not be easily determined whether 1) the gene or gene product is not functioning properly in the screening assay or 2) the gene or gene product is directly or indirectly involved in the biological process being assayed by the screening system.

Consequently, there is a need to provide methods and compositions for rapidly isolating portions of genomes associated with a known biological process and to screen such portions of genomes for activity without the necessity of transferring the gene of interest into an additional screening system.

### **Brief Description of the Figures**

**FIG. 1** shows a comparison between an application of a prior art reporter gene with methods described herein, and one embodiment of the invention. The prior art uses the beta-gal reporter and requires the establishment of clones prior to expression analysis.

One embodiment of this invention allows for the rapid identification of living cell clones from large multiclonal populations of **BLEC** (beta-lactamase expression construct) integrated cells. This is a significant advancement over the prior art, which requires the analysis of individual clones followed by the retrieving of selected clone from a duplicate clonal stock of living cells.

**FIG. 2** shows a representation of how one embodiment of the invention reports the expression of a pathway within a cell and can be used for screening.

**FIG. 3A** and **3B** shows a schematic plasmid map of **BLEC-1** and a viral vector map of **BLEC-RV1**, respectively.

**FIG. 4** shows the FACS analysis of a population of genomically **BLEC** integrated clones. Individual cells are plotted by fluorescent emission properties at 400 nm excitation. The x-axis represents green emission (530 nm). The y-axis represents blue emission (465 nm). Cells with a high blue/green ration will appear blue in color and cells with a low blue/green ratio will appear green in color. **A)** Unselected multiclonal population of **BLEC** integrated RBL-1 cell clones. **B)** Population of clones sorted from **3A (R1)** that were cultured for an additional 7 days and resorted. **C)** Population from **3B** with addition of 1microM ionomycin for 12 hours prior to sorting.

### Summary

The present invention recognizes that reporter genes, such as beta-lactamase polynucleotides, can be effectively used in living eukaryotic cells to functionally identify active portions of a genome directly or indirectly associated with a biological process. The present invention also recognizes for the first time that beta-lactamase activity can be measured using membrane permeant substrates in living cells incubated with a test chemical that directly or indirectly interacts with a portion of the genome having an integrated beta-lactamase polynucleotide. The present invention thus permits the rapid identification and isolation of genomic polynucleotides indirectly or directly associated with a defined biological process and identification of compounds that modulate such processes and regions of the genome. Because the identification of active genomic polynucleotides is permitted in living cells, further functional characterization can be

conducted using the same cells, and optionally, the same screening assay. The ability to functionally screen cells immediately after the rapid identification of a functionally active portion of a genome, without the necessity of transferring the identified portion of the genome into a secondary screening system, represents, among other things, a distinct  
5 advantage over an application of a prior art reporter gene with the methods described herein as described in **FIG. 1**.

The invention provides for a method of identifying portions of a genome, e.g. genomic polynucleotides, in a living cell using a polynucleotide encoding a protein with reporter gene activity, such as beta-lactamase activity, that can be detected with a  
10 membrane permeant substrate. Typically, the method involves inserting a polynucleotide encoding a protein with reporter gene activity into the genome of an organism using any method known in the art, developed in the future or described herein. Usually, a reporter gene expression construct will be used into integrate a reporter gene polynucleotide into a eukaryotic genome, as described herein. The cell, such as a eukaryotic cell, is usually  
15 contacted with a predetermined concentration of a modulator, either before or after integration of the reporter gene polynucleotide into the genome of the cell. Reporter gene activity is usually then measured inside the living cell, preferably with fluorescent, membrane permeant substrates that are transformed by the cell into membrane impermeant substrates as described herein.

The invention also provides for a method of identifying proteins or compounds that directly or indirectly modulate a genomic polynucleotide. Generally, the method comprises inserting a beta-lactamase expression construct into an eukaryotic genome, usually non-yeast, contained in at least one living cell, contacting the cell with a  
20 predetermined concentration of a modulator, and detecting beta-lactamase activity in the  
25 cell.

The invention also provides for a method of screening compounds with an active genomic polynucleotide that comprises: 1) optionally contacting a multiclonal population of cells with a first test chemical prior to separating said cells by a FACS, 2) separating by a FACS said multiclonal population of cells into reporter gene expressing cells and  
30 non-reporter gene expressing cells, wherein said reporter gene expressing cells have a detectable difference in cellular fluorescence properties compared to non-beta-lactamase

expressing cells. 3) contacting either population of cells with the same or a different test chemical. and 4) optionally repeating step (2), wherein said multi-clonal population of cells comprises eukaryotic cells having a beta-lactamase expression construct integrated into a genome of said cells and a membrane permanent beta-lactamase substrate transformed inside said cells to a membrane impermeant beta-lactamase substrate. The steps of this method can be repeated to permit additional characterization of identified clones.

The invention also includes powerful methods and compositions for identifying physiologically relevant cellular pathways and proteins of interest of known, unknown or partially known function. As shown in **FIG. 2** a cellular pathway may have more than one major intracellular signal. Two major intracellular pathways are shown ("A" and "B"). Each intracellular signal pathway may also have multiple branches. Each arm is shown as having three signaling pathways (A1, A2, and A3; and B1, B2, and B3). By generating a library of clones with a beta-lactamase expression construct, genomic polynucleotides for each signal pathway can be tagged or reported by the expression of beta-lactamase. Pathways not effected by the modulator (shown as C1, C2, and C3) are also tagged with beta-lactamase expression construct. Because the modulator only modulates the expression of pathways A1, A2, A3, B1, B2, and B3, only clones corresponding to these genomic integration sites are identified as being responsive to the modulator. Clones corresponding to sites C1, C2, and C3 remain unaltered and are not responsive to the modulator. Any individual, modulated clone can be immediately isolated, if not already isolated, and used for a drug discovery assay to screen test chemicals for activity for modulating the reported pathway, as described herein. Such methods and other aspects of the invention can be applied to other reporter genes.

The invention also includes tools for pathway identification and drug discovery that can be applied to a number of targets of interest and therapeutic areas including, proteins of interest, physiological responses even in the absence of a definitive target (e.g. immune response, signal transduction, neuronal function and endocrine function), viral targets, and orphan proteins.

Another aspect of the invention includes retroviral vectors and adeno-associated vectors that include a reporter gene. The reporter gene, once integrated into a genome, is

under the expression control of the genome. Such vectors can be used to identify genes and promoters as described herein.

## Detailed Description of the Invention

### 5 Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Generally, the nomenclature used herein, and the laboratory  
10 procedures in cell culture, molecular genetics, and nucleic acid chemistry and hybridization described below, are those well known and commonly employed in the art. Standard techniques are used for recombinant nucleic acid methods, polynucleotide synthesis, and microbial culture and transformation (e.g., electroporation, and lipofection). Generally, enzymatic reactions and purification steps are performed  
15 according to the manufacturer's specifications. The techniques and procedures are generally performed according to conventional methods in the art and various general references (see generally, Sambrook et al. Molecular Cloning: A Laboratory Manual, 2d ed. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., which is incorporated herein by reference) which are provided throughout this document. The  
20 nomenclature used herein, and the laboratory procedures in analytical chemistry, organic synthetic chemistry, and pharmaceutical formulation described below, are those well known and commonly employed in the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical formulation and delivery, and treatment of patients. As employed throughout the disclosure, the following terms, unless otherwise  
25 indicated, shall be understood to have the following meanings:

"Fluorescent donor moiety" refers to a fluorogenic compound or part of a compound (including a radical) which can absorb energy and is capable of transferring the energy to another fluorogenic molecule or part of a compound. Suitable donor fluorogenic molecules include, but are not limited to, coumarins and related dyes  
30 xanthene dyes such as fluoresceins, rhodols, and rhodamines, resorufins, cyanine dyes, bimanes, acridines, isoindoles, dansyl dyes, aminophthalic hydrazides such as luminol

and isoluminol derivatives, aminophthalimides, aminonaphthalimides, aminobenzofurans, aminoquinolines, dicyanohydroquinones, and europium and terbium complexes and related compounds.

"Quencher" refers to a chromophoric molecule or part of a compound that is capable of reducing the emission from a fluorescent donor when attached to the donor. Quenching may occur by any of several mechanisms including fluorescence resonance energy transfer, photoinduced electron transfer, paramagnetic enhancement of intersystem crossing, Dexter exchange coupling, and excitation coupling such as the formation of dark complexes.

"Acceptor" refers to a quencher that operates via fluorescence resonance energy transfer. Many acceptors can re-emit the transferred energy as fluorescence. Examples include coumarins and related fluorophores, xanthenes such as fluoresceins, rhodols, and rhodamines, resorufins, cyanines, difluoroboradiazaindacenes, and phthalocyanines. Other chemical classes of acceptors generally do not re-emit the transferred energy. Examples include indigos, benzoquinones, anthraquinones, azo compounds, nitro compounds, indoanilines, and di- and tri-phenylmethanes.

"Dye" refers to a molecule or part of a compound that absorbs specific frequencies of light, including but not limited to ultraviolet light. The terms "dye" and "chromophore" are synonymous.

"Fluorophore" refers to a chromophore that fluoresces.

"Membrane-permeant derivative" refers a chemical derivative of a compound of that increases membrane permeability of the compound. These derivatives are made better able to cross cell membranes, i.e. membrane permeant, because hydrophilic groups are masked to provide more hydrophobic derivatives. Also, the masking groups are designed to be cleaved from the fluorogenic substrate within the cell to generate the derived substrate intracellularly. Because the substrate is more hydrophilic than the membrane permeant derivative it becomes trapped within the cell.

"Isolated polynucleotide" refers to a polynucleotide of genomic, cDNA, or synthetic origin or some combination there of, which by virtue of its origin, the "isolated polynucleotide" (1) is not associated with the cell in which the "isolated polynucleotide"

is found in nature, or (2) is operably linked to a polynucleotide which it is not linked to in nature.

"Isolated protein" refers to a protein of cDNA, recombinant RNA, or synthetic origin, or some combination thereof, which by virtue of its origin the "isolated protein"

5 (1) is not associated with proteins found as it is normally found with in nature, or (2) is isolated from the cell in which it normally occurs, or (3) is isolated free of other proteins from the same cellular source, e.g. free of human proteins, or (4) is expressed by a cell from a different species, or (5) does not occur in nature.

"Polypeptide" as used herein as a generic term to refer to native protein,

10 fragments, or analogs of a polypeptide sequence. Hence, native protein, fragments, and analogs are species of the polypeptide genus. Preferred, beta-lactamase polypeptides include those with the polypeptide sequence represented in the **SEQUENCE ID. LISTING** and any other polypeptide or protein having similar beta-lactamase activity as measured by one or more of the assays described herein. beta-lactamase polypeptide or

15 proteins can include any protein having sufficient activity for detection in the assays described herein.

"Naturally-occurring" as used herein, as applied to an object, refers to the fact that an object can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in

20 nature and which has not been intentionally modified by man in the laboratory is naturally-occurring.

"Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression

25 of the coding sequence is achieved under conditions compatible with the control sequences.

"Control sequence" refers to polynucleotide sequences which are necessary to effect the expression of coding and non-coding sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism; in

30 prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence; in eukaryotes, generally, such control sequences



include promoters and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, components whose presence can influence expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

5 "Polynucleotide" refers to a polymeric form of nucleotides of at least ten bases in length, either ribonucleotides or deoxynucleotides or a modified form of either type of nucleotide. The term includes single and double stranded forms of DNA. "Genomic polynucleotide" refers to a portion of a genome. "Active genomic polynucleotide" or "active portion of a genome" refer to regions of a genome that can be up regulated,  
10 down-regulated or both, either directly or indirectly, by a biological process. "Directly," in the context of a biological process or processes, refers to direct causation of a process that does not require intermediate steps, usually caused by one molecule contacting or binding to another molecule (the same type or different type of molecule). For example, molecule A contacts molecule B, which causes molecule B to exert effect X that is part of  
15 a biological process. "Indirectly," in the context of a biological process or processes, refers to indirect causation that requires intermediate steps, usually caused by two or more direct steps. For example, molecule A contacts molecule B to exert effect X which in turn causes effect Y.

"Beta-lactamase polynucleotide" refers to a polynucleotide encoding a protein  
20 with beta-lactamase activity. Preferably, the protein with beta-lactamase activity can be measured in a FACS at about 22°degrees using a CCF2/AM beta-lactamase substrate at a level of about 1,000 such protein molecules or less per cell. More preferably, the protein with beta-lactamase activity can measured be in a FACS at about 22° degrees using a CCF2/AM beta-lactamase substrate at a level of about 300 to 1,000 such protein  
25 molecules per cell. More preferably, the protein with beta-lactamase activity can measured be in a FACS at about 22° degrees using a CCF2/AM beta-lactamase substrate at a level of about 25 to 300 such protein molecules per cell. Proteins with beta-lactamase activity that require more than 1,000 molecules of such protein per cell for detection with a FACS at about 22° degrees using a CCF2/AM beta-lactamase substrate  
30 can be used and preferably have at least about 5% of the activity of the protein with SEQ. ID. NO.: 1.

“Reporter gene” means a gene that encodes a reporter, such as are known in the art or are later developed. Reporter genes can encode enzymes such as beta-lactamase, beta-galactosidase, and luciferase (for beta-lactamase, see WO 96/30540 to Tsien, published October 3, 1996). Reporter genes can also encode fluorescent proteins, such as green fluorescent protein (GFP) or mutants thereof as they are known in the art or are later developed (see, U.S. Patent No. 5,625,048. to Tsien, issued April 29, 1997; WO 96/23810 to Tsien, published August 8, 1996; WO 97/28261 to Tsien, published August 7, 1997; and PCT/US97/12410 to Tsien, filed July 16, 1996) . The products of reporter genes can be detected using methods known in the art, such as the use of chromogenic or fluorogenic substrates for enzymes. Chromogenic or fluorogenic readouts can be detected using, for example, optical methods such as absorbance or fluorescence. A reporter gene can be part of a reporter gene construct, such as a plasmid or viral vector, such as a retrovirus or adeno-associated virus.

“Sequence homology” refers to the proportion of base matches between two nucleic acid sequences or the proportion amino acid matches between two amino acid sequences. When sequence homology is expressed as a percentage, e.g., 50%, the percentage denotes the proportion of matches over the length of sequence from a desired sequence (e.g. beta-lactamase sequences, such as SEQ. ID. NO.: 1) that is compared to some other sequence. Gaps (in either of the two sequences) are permitted to maximize matching; gap lengths of fifteen bases or less are usually used, six bases or less are preferred with two bases or less more preferred. When using oligonucleotides as probes or treatments the sequence homology between the target nucleic acid and the oligonucleotide sequence is generally not less than seventeen target base matches out of twenty possible oligonucleotide base pair matches (85%); preferably not less than nine matches out of ten possible base pair matches (90%), and most preferably not less than 19 matches out of 20 possible base pair matches (95%).

“Selectively hybridize” refers to detectably and specifically bind. Polynucleotides, oligonucleotides and fragments thereof selectively hybridize to target nucleic acid strands, under hybridization and wash conditions that minimize appreciable amounts of detectable binding to nonspecific nucleic acids. High stringency conditions can be used to achieve selective hybridization as is known in the art and discussed herein.

Generally, the nucleic acid sequence homology between the polynucleotides, oligonucleotides, and fragments thereof and a nucleic acid sequence of interest will be at least 30%, and, more typically, with preferably increasing homologies of at least about 40%, 50%, 60%, 70%, and 90%.

Typically, hybridization and washing conditions are performed at high stringency according to conventional hybridization procedures. Positive clones are isolated and sequenced. For illustration and not for limitation, a full-length polynucleotide corresponding to the nucleic acid sequence of **SEQ. ID.NO. 1** may be labeled and used as a hybridization probe to isolate genomic clones from a the appropriate target library in  $\lambda$ EMBL4 or  $\lambda$ GEM11 (Promega Corporation, Madison, Wisconsin); typical hybridization conditions for screening plaque lifts (Benton and Davis (1978) Science 196: 180) can be: 50% formamide, 5 x SSC or SSPE, 1 to 5 x Denhardt's solution, 0.1 to 1% SDS, 100-200  $\mu$ g sheared heterologous DNA or tRNA, 0-10% dextran sulfate,  $1 \times 10^5$  to  $1 \times 10^7$  cpm/ml of denatured probe with a specific activity of about  $1 \times 10^8$  cpm/ $\mu$ g, and incubation at about 42°C for about 6 to 36 hours. Prehybridization conditions are essentially identical except that probe is not included and incubation time is typically reduced. Washing conditions are typically 1 to 3 x SSC, 0.1 to 1% SDS, 50 to 70°C with change of wash solution at about 5 to 30 minutes. Cognate sequences, including allelic sequences, can be obtained in this manner.

Two amino acid sequences are homologous if there is a partial or complete identity between their sequences. For example, 85% homology means that 85% of the amino acids are identical when the two sequences are aligned for maximum matching. Gaps (in either of the two sequences being matched) are allowed in maximizing matching, gap lengths of five or less are preferred with two or less being more preferred. Alternatively, and preferably, two protein sequences (or polypeptide sequences derived from them of at least 30 amino acids in length) are homologous, as this term is used herein, if they have an alignment score of at more than five (in standard deviation units) using the program ALIGN with the mutation data matrix and a gap penalty of 6 or greater. See Dayhoff, M.O., in Atlas of Protein Sequence and Structure, 1972, volume 5, National Biomedical Research Foundation, pp. 101-110, and Supplement 2 to this volume, pp. 1-10. The two sequences, or parts thereof, are more preferably homologous

if their amino acids are greater than or equal to 30% identity when optimally aligned using the ALIGN program.

"Corresponds to" refers to a polynucleotide sequence is homologous (i.e., is identical, not strictly evolutionarily related) to all or a portion of a reference polynucleotide sequence, or that a polypeptide sequence is identical to all or a portion of a reference polypeptide sequence. In contradistinction, the term "complementary to" is used herein to mean that the complementary sequence is homologous to all or a portion of a reference polynucleotide sequence. For illustration, the nucleotide sequence "TATAC" corresponds to a reference sequence "TATAC" and is complementary to a reference sequence "GTATA".

The following terms are used to describe the sequence relationships between two or more polynucleotides: "reference sequence," "comparison window," "sequence identity," "percentage of sequence identity," and "substantial identity." A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length cDNA or gene sequence given in a sequence listing such as a **SEQ. ID. NO.: 1**, or may comprise a complete cDNA or gene sequence. Generally, a reference sequence is at least 20 nucleotides in length, frequently at least 25 nucleotides in length, and often at least 50 nucleotides in length. Since two polynucleotides may each (1) comprise a sequence (i.e., a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) may further comprise a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity. A "comparison window", as used herein, refers to a conceptual segment of at least 20 contiguous nucleotide positions wherein a polynucleotide sequence may be compared to a reference sequence of at least 20 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by the local

homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2: 482, by the homology alignment algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48: 443, by the search for similarity method of Pearson and Lipman (1988) Proc. Natl. Acad. Sci. (U.S.A.) 85: 2444, by computerized implementations of these algorithms (GAP,

5 BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0. Genetics Computer Group, 575 Science Dr., Madison, WI), or by inspection, and the best alignment (i.e., resulting in the highest percentage of homology over the comparison window) generated by the various methods is selected. The term "sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide  
10 basis) over the window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in  
15 the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The terms "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 30 percent sequence identity, preferably at least 50 to 60 percent sequence identity, more usually at least 60 percent sequence identity as  
20 compared to a reference sequence over a comparison window of at least 20 nucleotide positions, frequently over a window of at least 25-50 nucleotides, wherein the percentage of sequence identity is calculated by comparing the reference sequence to the polynucleotide sequence which may include deletions or additions which total 20 percent or less of the reference sequence over the window of comparison.

25 As applied to polypeptides, the term "substantial identity" means that two peptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least 30 percent sequence identity, preferably at least 40 percent sequence identity, more preferably at least 50 percent sequence identity, and most preferably at least 60 percent sequence identity. Preferably, residue positions, which are  
30 not identical, differ by conservative amino acid substitutions. Conservative amino acid substitutions refer to the interchangeability of residues having similar side chains. For

example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine. A group of amino acids having aliphatic-hydroxyl side chains is serine and threonine. A group of amino acids having amide-containing side chains is asparagine and glutamine. A group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan. A group of amino acids having basic side chains is lysine, arginine, and histidine. A group of amino acids having sulfur-containing side chains is cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, glutamic-aspartic, and asparagine-glutamine.

"Polypeptide fragment" refers to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion, but where the remaining amino acid sequence is usually identical to the corresponding positions in the naturally-occurring sequence deduced, for example, from a full-length cDNA sequence (e.g., the sequence shown in **SEQ. ID. NO.: 1**). "beta-lactamase polypeptides fragment" refers to a polypeptide that is comprised of a segment of at least 25 amino acids that has substantial identity to a portion of the deduced amino acid sequence shown in **SEQ. ID. NO.:1** and which has at least one of the following properties: (1) specific binding to a beta-lactamase substrate, preferably cephalosporin, under suitable binding conditions, or (2) the ability to effectuate enzymatic activity, preferably cephalosporin backbone cleavage activity, when expressed in a mammalian cell. Typically, analog polypeptides comprise a conservative amino acid substitution (or addition or deletion) with respect to the naturally occurring sequence. Analogs typically are at least 300 amino acids long, preferably at least 500 amino acids long or longer, most usually being as long as full-length naturally-occurring polypeptide.

"Modulation " refers to the capacity to either enhance or inhibit a functional property of a biological activity or process (e.g., enzyme activity or receptor binding). Such enhancement or inhibition may be contingent on the occurrence of a specific event, such as activation of a signal transduction pathway, and/or may be manifest only in particular cell types.

The term "modulator" refers to a chemical (naturally occurring or non-naturally occurring), such as a biological macromolecule (e.g. nucleic acid, protein, non-peptide, or organic molecule), or an extract made from biological materials such as bacteria, plants,

fungi, or animal (particularly mammalian) cells or tissues. Modulators are typically evaluated for potential activity as inhibitors or activators (directly or indirectly) of a biological process or processes (e.g., agonist, partial antagonist, partial agonist, antagonist, antineoplastic agents, cytotoxic agents, inhibitors of neoplastic transformation or cell proliferation, cell proliferation-promoting agents, and the like) by inclusion in assays described herein. The activity of a modulator may be known, unknown or partial known.

The term "test chemical" refers to a chemical to be tested by one or more method(s) of the invention as a putative modulator. A test chemical is usually not known to bind to the target of interest. The term "control test chemical" refers to a chemical known to bind to the target (e.g., a known agonist, antagonist, partial agonist or inverse agonist). The term "test chemical" does not typically include a chemical added as a control condition that alters the function of the target to determine signal specificity in an assay. Such control chemicals or conditions include chemicals that 1) non-specifically or substantially disrupt protein structure (e.g., denaturing agents (e.g., urea or guanidium), chaotropic agents, sulfhydryl reagents (e.g., dithiotritol and beta-mercaptoethanol), and proteases), 2) generally inhibit cell metabolism (e.g., mitochondrial uncouplers) and 3) non-specifically disrupt electrostatic or hydrophobic interactions of a protein (e.g., high salt concentrations, or detergents at concentrations sufficient to non-specifically disrupt hydrophobic interactions). The term "test chemical" also does not typically include chemicals known to be unsuitable for a therapeutic use for a particular indication due to toxicity of the subject. Usually, various predetermined concentrations test chemicals are used for screening such as 0.01 microM, 0.1 microM, 1.0 microM, and 10.0 microM.

The term "target" refers to a biochemical entity involved a biological process. Targets are typically proteins that play a useful role in the physiology or biology of an organism. A therapeutic chemical binds to target to alter or modulate its function. As used herein, targets can include cell surface receptors, G-proteins, kinases, ion channels, phospholipases and other proteins mentioned herein.

The terms "label" or "labeled" refers to incorporation of a detectable marker, e.g., by incorporation of a radiolabeled amino acid or attachment to a polypeptide of biotinyl moieties that can be detected by marked avidin (e.g., streptavidin containing a fluorescent

marker or enzymatic activity that can be detected by optical or colorimetric methods).

Various methods of labeling polypeptides and glycoproteins are known in the art and may be used. Examples of labels (e.g. for polypeptides or polynucleotides) include, but

are not limited to, the following: radioisotopes (e.g.,  $^3\text{H}$ ,  $^{14}\text{C}$ ,  $^{35}\text{S}$ ,  $^{125}\text{I}$ ,  $^{131}\text{I}$ ), fluorescent

5 labels (e.g., FITC, rhodamine, and lanthanide phosphors), enzymatic labels (or reporter genes) (e.g., enzymatic reporter genes horseradish peroxidase, beta-galactosidase,

luciferase and alkaline phosphatase; and non-enzymatic reporter genes (e.g., fluorescent proteins)), chemiluminescent, biotinyl groups, predetermined polypeptide epitopes

recognized by a secondary reporter (e.g., leucine zipper pair sequences, binding sites for

10 secondary antibodies, metal binding domains, and epitope tags). "Substantially pure"

refers to an object species is the predominant species present (i.e., on a molar basis it is more abundant than any other individual species in the composition), and preferably a

substantially purified fraction is a composition wherein the object species comprises at least about 50 percent (on a molar basis) of all macromolecular species present.

15 Generally, a substantially pure composition will comprise more than about 80 percent of

all macromolecular species present in the composition, more preferably more than about 85%, 90%, 95%, and 99%. Most preferably, the object species is purified to essential

homogeneity (contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists essentially of a single

20 macromolecular species.

"Pharmaceutical agent or drug" refers to a chemical or composition capable of inducing a desired therapeutic effect when properly administered (e.g. using the proper amount and delivery modality) to a patient.

Other chemistry terms herein are used according to conventional usage in the art,

25 as exemplified by The McGraw-Hill Dictionary of Chemical Terms (ed. Parker, S.,

1985), McGraw-Hill, San Francisco, incorporated herein by reference).

## **Introduction**

30 The present invention recognizes that reporter genes, such as beta-lactamase polynucleotides, can be effectively used in living eukaryotic cells to functionally identify



active portions of a genome directly or indirectly associated with a biological process.

The present invention also recognizes for the first time that reporter gene activity, such as beta-lactamase activity, can be measured using membrane permeant substrates in living cells incubated with a test chemical that directly or indirectly interacts with a portion of the genome having an integrated reporter gene. The present invention, thus, permits the rapid identification and isolation of genomic polynucleotides indirectly or directly associated with a defined biological process and identification of compounds that modulate such processes and regions of the genome. Because the identification of active genomic polynucleotides is permitted in living cells, further functional characterization can be conducted using the same cells, and, optionally, the same screening assay. The ability to functionally screen immediately after the rapid identification of a functionally active portion of a genome, without the necessity of transferring the identified portion of the genome into a secondary screening system, represents, among other things, a distinct advantage over an application of a prior art reporter gene and methods as shown in **FIG.**

1.

As a non-limiting introduction to the breadth of the invention, the invention includes several general and useful aspects, including:

- 1) a method for identifying genes or gene products directly or indirectly associated (e.g. regulated) with a biological process of interest (that can be modulated by a compound) using a genomic polynucleotide operably linked to a polynucleotide encoding a protein with beta-lactamase activity or other reporter gene,
- 2) a method for identifying proteins (e.g. orphan proteins or known proteins) or compounds that directly or indirectly modulate (e.g. activate or inhibit transcription) a genomic polynucleotide operably linked to a polynucleotide encoding a protein with beta-lactamase activity,
- 3) a method of screening for an active genomic polynucleotide (e.g. enhancer, promoter or coding region in the genome) that can be directly or indirectly associated (e.g. regulated) with a biological process of interest (that can be modulated by a compound) using a genomic

polynucleotide operably linked to a polynucleotide encoding a protein with beta-lactamase activity that can be detected by FACS using a fluorescent, membrane permeant beta-lactamase substrate,

- 4) eukaryotic cells with a genomic polynucleotide operably linked to a polynucleotide encoding a protein with beta-lactamase activity, and
- 5) polynucleotides (including vectors) related to the above methods and cells.

These aspects of the invention, as well as others described herein, can be achieved by using the methods and compositions of matter described herein. To gain a full appreciation of the scope of the invention, it will be further recognized that various aspects of the invention can be combined to make desirable embodiments of the invention. For example, the invention includes a method of identifying compounds that modulate active genomic polynucleotides operably linked to a protein with beta-lactamase activity that can be detected by FACS using a fluorescent, membrane permeant beta-lactamase substrate. Such combinations result in particularly useful and robust embodiments of the invention.

### **Methods for Rapidly Identifying Functional Portions of a Genome**

The invention provides for a method of identifying portions of a genome, e.g. genomic polynucleotides, in a living cell using a polynucleotide encoding a reporter gene, such as a beta-lactamase activity, that can be detected with a membrane permeant substrate. Preferably, the method involves inserting a polynucleotide encoding a protein with beta-lactamase activity into the genome of an organism using any method known in the art, developed in the future or described herein. Usually, a reporter gene expression construct will be used into integrate a reporter gene into a eukaryotic genome, as described herein. The cell, such as a eukaryotic cell, is usually contacted with a predetermined concentration of a modulator, either before or after integration of the reporter gene. Reporter gene activity, such as beta-lactamase activity, is usually then measured inside the living cell, preferably with fluorescent, membrane permeant substrates that are transformed by the cell into membrane impermeant substrates as

described herein and PCT Publication No. WO96/30540, published October 3, 1996, by Tsien et al.

Once reporter genes are integrated into the genome of interest, they become under the transcriptional control of the genome of the host cell. Integration into the genome is usually stable, as described herein and known in the art. Transcriptional control of the genome often results from receptor (e.g. intracellular or cell surface receptor) activation, which can regulate transcriptional and translational events to change the amount of protein present in the cell. The amount of protein present with  $\beta$ -lactamase activity can be measured via its enzymatic action on a substrate. Normally, the substrate is a small uncharged molecule that, when added to the extracellular solution, can penetrate the plasma membrane to encounter the enzyme. A charged molecule can also be employed, but the charges are generally masked by groups that will be cleaved by endogenous or heterologous cellular enzymes or processes (e.g., esters cleaved by cytoplasmic esterases). As described more fully herein and in PCT Publication No. WO96/30540 published October 3, 1996, by Tsien et al., which is herein incorporated by reference, the use of substrates that exhibit changes in their fluorescence spectra upon interaction with an enzyme are particularly desirable. In some assays, the fluorogenic substrate is converted to a fluorescent product by beta-lactamase activity. Alternatively, the fluorescent substrate changes fluorescence properties upon conversion by beta-lactamase activity. Preferably, the product should be very fluorescent to obtain a maximal signal, and very polar, to stay trapped inside the cell.

### *Vectors and Integration*

Vectors, such as viral and plasmid vectors, can be used to introduce genes or genetic material of the invention into cells, preferably by integration into the host cell genome. Such viral vectors can be any appropriate viruses, such as retroviruses, adenoviruses, adeno-associated viruses, papillomaviruses, herpes viruses, or any ecotropic or amphotropic virus, preferably a retrovirus. The viruses can be, for example, retroviruses or any other virus that are replicatively competent or modified to be replicatively deficient, cytomegalovirus, Friend leukemia virus, myeloproliferative

sarcoma virus, SL3-3, SIV, HIV, Rouse Sarcoma Virus, or Moloney virus such as Moloney murine leukemia virus. Such viral vectors can be DNA or RNA based viruses. Examples of DNA viral vectors include adenoviral, adeno-associated viral, papilloma viral, herpes viral, Epstein Barr viral, or SV40 viral vectors. Examples of RNA viral vectors include alphaviral (e.g. Sindbis and Semliki Forest Virus), and retroviral (e.g. including lentiviral vectors such as HIV and SIV, as well as Murine oncoviruses such as Moloney Murine Leukemia Virus, Moloney Murine Sarcoma Virus, SL3-3, Rous Sarcoma Virus, Cytomegalovirus and derivatives thereof). The retroviruses can be pseudotyped to contain envelopes with various host ranges including murine amphotropic (e.g. 4070A for PA317, AM12, and FLYA 13 packaging cells), murine ecotropic (for example GP+E86 packaging cells), GALV (from gibbon ape leukemia virus; for example PG13 packaging cells), FeLV (Feline leukemia virus; for example FLYRD18 packaging cells). Preferably, retroviral vectors or adeno-associated viral vectors are used. Typically, the viruses are replicatively deficient, but do not need to be so to be useful in the present invention. General types of such viral vectors are known in the art (see, U.S. Patent No. 5,627,058 to Ruley et al. issued May 6, 1997; U.S. Patent No. 5,364,783 to Ruley et al., issued November 15, 1994; U.S. Patent No. 5,399,346 issued to Anderson et al. on March 21, 1995; Bandara et al., DNA and Cell Biology 11:227-231 (1992); Berkner, BioTechniques 6:616-629 (1989); U.S. Patent No. 5,240,846 issued to Collins et al. on August 31, 1993; Culver and Blaese, TIG 10:171-178 (1994); Goldman et al., Gene Therapy, 3:811-818 (1996); Holmberg et al., J. Liposome Res. 1:393-406 (1990); Karlsson et al., The EMBO J. 5:2377-2385 (1986); Krul et al., Cancer Immunol. Immunother. 43:44-48 (1996); Larrick and Burck, Gene Therapy Application of Molecular Biology, Elsevier, New York (1991); Mountford and Smith, supra (1995); Mountford et al., supra, (1994); Fukushige and Sauer, supra, (1992); Shapiro and Senapathy, supra, (1987); Niwa et al., J. Biochem., 113:343-349 (1993); Wurst et al., supra (1995); Reddy et al., supra, (1992); Friedrich and Soriano, Methods in Enzymology 225:681-701, (1991); Gossler et al., supra (1989); Friedrich and Soriano, Genes and Development, 5:1513-1523 (1991); Hill and Wurst, supra, (1993); Skarnes et al., supra, (1992)).

A vector of the present invention can comprise a nucleic acid sequence encoding a reporter gene, a splice acceptor sequence, and a splice donor sequence. The splice acceptor sequences can be those known in the art or later identified, such as engrailed-2 (en-2) splice acceptor. Splice donor sequences can be those known in the art or later identified, such as SV40 or beta-actin splice donor. Such vectors can be used for integration into a genome to identify promoters and genes using the methods of the present invention. Preferably, the splice acceptor sequence and the splice donor sequence flank the reporter gene (e.g. splice acceptor sequence, reporter gene, and splice donor sequence). The reporter gene can encode, for example, a beta-lactamase, a luciferase, a green fluorescent protein (GFP), beta-galactosidase, or other reporter gene as that term is understood in the art, including cell surface markers, such as CD<sub>4</sub> or the truncated nerve growth factor (NGFR) (for GFP, see WO 96/23810 to Tsien, published 8/8/96; Heim et al., Current Biology, 2:178-182 (1996), Heim et al., Proc. Natl. Acad. Sci. USA (1995), or Heim et al., Science 373:663-664 (1995), for beta-lactamase, see WO 96/30540 to Tsien published 11/3/96).

A vector of the present invention can comprise more than one such reporter gene, as well as a selectable marker. For example, the vector can include two detectable reporter genes or two selectable markers, or one detectable reporter gene and one selectable marker. Typically, such reporter genes or selectable markers are flanked by the splice acceptor or donor. Preferred examples include nucleic acid sequences that encode beta-lactamase and GFP or beta-lactamase and neomycin resistance. The vector can also include a fusion protein wherein said fusion protein can comprise more than a reporter gene and a selectable marker. Preferred examples include beta-lactamase-neomycin resistance fusion protein or beta-lactamase-puromycin fusion protein

The reporter gene can also comprise an ATG sequence in the 5' region of the reporter gene to enhance or initiate translation of a reporter gene (see, for example, Friedrich and Soriano, Genes & Development 5:1513-1523 (1991); and Cavener et al., Nucleic Acids Res, 19:3185-3192 (1991)). The region around the ATG sequence can be optimized for translation in mammalian cells using, for example, a Kozak's sequence (see, Kozak, Nucleic Acids Res. 15:8125-8148 (1987)). The region 5' of the reporter gene can also be operably linked to an internal ribosome entry site (IRES) to reduce the

need for in-frame insertion of the reporter gene into the proper reading frame of an endogenous gene while practicing a method of the present invention and increase the expression of the reporter gene several fold (see, for example Mountford and Smith, TIG 11:179-184 (1995), and Mountford et al., Proc. Natl. Acad. Sci. USA 91:4303-4307 (1994), each of which are incorporated by reference). The reporter gene can also comprise a poly-adenylation site at its 3' end to enhance the expression of the product of the reporter gene by stabilizing RNA molecules (see, for example, Freidrich and Soriano, Genes & Development 5:1513-1523 (1991)).

A vector of the present invention can comprise 5' and/or 3' long terminal repeat regions (LTRs) or deleted LTRs (dLTRs) (see, Coffin et al., Retroviruses, Cold Spring Harbor Laboratory Press, N.Y. (1996); Ausubel et al., Current Protocols in Molecular Biology, John Wiley & Sons (1994); Miller et al., BioTechniques 7:980-990 (1989); Vile et al., Brit. Med. Bull. 51:12-30 (1995); and Yu et al., Proc. Natl. Acad. Sci. USA 83:3194-3198 (1986) in order to aid the integration of the vector into the genome of the host cell (see, Freidrich and Soriano, Genes & Development 5:1513-1523 (1991); Chaulika et al., J. Virol. 70:1792-1798 (1996); Mayo et al., Blood 86:3139-3150 (1995); Wybier-Franqui et al. AIDS Res. Hum. Retroviruses 11:829-836 (1995); Miyazawa et al., J. Vet. Med. Sci. 56:869-872 (1994); and Miyazawa et al., Arch. Virol. 139:37-48 (1994)). The LTRs preferably flank the vector constructs discussed above. Furthermore, the components of the vector described above can be provided in a forward or reverse orientation to enhance packaging titer by eliminating poly-A signals in the forward orientation (see, Friedrich and Soriano, Genes and Development 5:1513-1524 (1991)). Furthermore, the present invention contemplates using double copy vectors, such as SIN vectors (see, Vile and Russell, British Medical Bulletin 51:12-30 (1995)). Furthermore, the vector can be modified to eliminate the retroviral splice donor sequence adjacent to the 5' LTR to accommodate splice acceptor sequences in the forward orientation relative to the retroviral transcript.

Vectors of the present invention can comprise a reporter gene with or without an upstream or downstream IRES sequence. Furthermore, a vector the present invention can comprise a eukaryotic promoter, such as they are known in the art or later

identified, such as CMV or actin. A vector of the present invention can also include an inducible promoter, such as tetracycline inducible promoters or others known in the art or later identified.

Vectors of the present invention, such as retroviral or adeno-associated viral, can encode an operable selective marker so that cells that have been transformed can be positively selected for. Such selective marker can be antibiotic resistance factors, such as neomycin resistance, such as *neo* and can be bleo (a fusion protein of beta-lactamase and neo), hygromycin resistance, puromycin resistance, and can also be cell surface markers, such nerve growth factor receptor or cytoplasmicly truncated versions thereof. Alternatively, cells can be negatively selected for using an enzyme, such as herpes simplex virus thymidine kinase (HSVTK) that converts a pro-toxin (gancyclovir) into a toxin.

Retroviral vectors of the present invention can be made using methods known in the art (see, Sambrook et al., supra, (1989)). For example, plasmids encoding elements of a retrovirus can be made using standard recombinant DNA methods. These plasmids are introduced into retroviral packaging cell lines, such as PT67, using standard gene transfer techniques, such as electroporation, calcium phosphate transfection, and lipofection. Packaging cell lines with integrated plasmid constructs, known as retroviral producer cells, can be selected by antibiotic resistance or cell sorting for a reporter gene, when appropriate. Ping-pong techniques can be used to increase the titer of the retroviral vectors (Kozak and Kabat, J. Virol. 64:3500-3508 (1990)). Identification of high titer producer cell clones can be accomplished using RNA dot blot hybridization, antibiotic resistance, or reporter gene expression. Titers of retrovirus preparations can be increased by culturing retroviral producer cells at 32°C rather than 37°C, selecting for packaging cell functions, and concentrating methods such as centrifugation to pellet retroviruses and by lyophilization. Also, transduction efficiency of retroviruses can be increased by centrifugation methods as are known in the art and by performing transductions at 32°C rather than 37°C. Virus titers can also be increased by co-cultivating producer cells with target cells and be incubating target cells in phosphate-free media prior to infection.

Viral vectors, such as retroviral vectors, are available that are suitable for these purposes, such as pSIR vector (available from ClonTech of California with PT67 packaging cells) GgU3Hisen and GgTNKneoU3 and GgTKNeoen variants of Moloney murine leukemia virus, are available. Vector modifications can be made that allow more efficient integration into the host cell genome. Such modifications include sequences that enhance integration or known methods to promote nucleic acid transportation into the nucleus of the host cell. Retroviral vectors, such as those described in U.S. Patent Number 5,364,783 to Ruley and von Melchner can also be used.

Preferable retroviral vectors include the configurations set forth below. In addition, all of the vectors can be provided with the 3' LTR and 5' LTR exchanged (for example, the insert is provided in a reversed orientation) and/or can have at least one LTR from a self-inactivating retrovirus, such as a dLTR. Furthermore, if present, an endogenous splice donor of the retroviral vector can be deleted or mutated to be non-functional.

- 5'LTR/splice acceptor/beta-lactamase/splice donor/LTR3'
- 3'LTR/splice acceptor/beta-lactamase/poly-A/LTR5'
- 3'LTR/splice acceptor/IRES/beta-lactamase/poly-A/LTR5'
- 3'LTR/splice acceptor/beta-lactamase/poly-A/beta-actin promoter/neo/splice donor/LTR5'
- 3'LTR/splice acceptor/beta-lactamase/poly-A/beta-actin promoter/neo/poly-A/LTR5'
- 3'LTR/splice acceptor/IRES/beta-lactamase/poly-A/beta-actin promoter/neo/splice donor/LTR5'
- 3'LTR/splice acceptor/IRES/beta-lactamase/poly-A/beta-actin promoter/neo/poly-A/LTR5'
- 3'dLTR/splice acceptor/beta-lactamase/poly-A/beta-actin promoter/neo/poly-A/dLTR5'
- 3'dLTR/splice acceptor/IRES/beta-lactamase/poly-A/beta-actin promoter/neo/poly-A/dLTR5'
- 5'LTR/splice acceptor/beta-lactamase/beta-actin promoter/neo/splice donor/dLTR3'
- 5'LTR/mutant splice donor/splice acceptor/beta-lactamase/actin promoter/neo/LTR3'
- 5'LTR/splice acceptor/beta-lactamase/beta-actin promoter/neo/dLTR3'
- 3'LTR/splice acceptor/IRES/beta-lactamase/poly-A/beta-actin promoter/neo/poly-A/LTR5'
- 3'LTR/splice acceptor/beta-lactamase/poly-A/beta actin promoter/neo/splice donor/dLTR5'
- 3'LTR/splice acceptor/IRES/beta-lactamase/poly-A/beta actin promoter/neo/splice donor/dLTR5'
- 5'LTR/splice acceptor/reporter gene/eukaryotic promoter/selectable marker/LTR 3'
- 5'LTR/splice acceptor/reporter gene/IRES/eukaryotic promoter/selectable marker/LTR3'
- 3'LTR/splice acceptor/reporter gene/poly-A/eukaryotic promoter/selectable marker/poly-A/LTR5'
- 3'LTR/splice acceptor, reporter gene/IRES/selectable marker/poly-A/LTR5'
- 3'LTR/splice acceptor/reporter gene/poly-A/eukaryotic promoter/reporter gene/splice



donor/LTR5'  
5'LTR/splice acceptor/reporter gene/LTR3'  
3'LTR/splice acceptor/reporter gene/splice donor/LTR5'

5 Additional retroviral vectors of the present invention include double copy  
retroviral vectors. These vectors can be used in the methods of the present invention to  
identify promoters or genes. These vectors are made using standard methods in  
molecular biology as discussed above (see, Sambrook et al., supra, 1989). In double  
copy retroviral vectors, the reporter gene is cloned into the U3 region of an LTR, such as  
10 the 3'LTR. When in an appropriate cell, reverse transcription can result in a duplication  
of the LTR region such that the reporter gene can be present in both the 3'LTR and the  
5'LTR upon integration into the genome of a cell. Preferred double copy retroviral  
vectors of the present invention include the following integrated into the U3 region of the  
3'LTR: a reporter gene is an optimal translation start, a splice acceptor followed by a  
15 reporter gene, or an IRES sequence followed by a reporter gene. These vectors are  
preferred for identifying promoters, but are also useful for identifying genes.

Vectors of the present invention can also be adeno-associated viruses (AAVs).  
These AAV vectors can be used in the methods of the present invention to identify  
promoters and genes. Generally, AAVs used to identify promoters contain a reporter  
20 gene with consensus translational initiation sequences, such as Kozak sequences.  
Generally, AAVs used to identify genes contain a reporter gene downstream of a splice  
acceptor site.

AAV vectors of the present invention can be made using standard recombinant  
DNA techniques (see, Sambrook et al, supra 1989). For example, AAV tagging  
25 constructs made using methods known in the art can be transfected into an appropriate  
packaging cell line (see, Walter and High, Advances in Veterinary Medicine, 40:119-134  
(1997); Linden et al. Proc. Natl. Acad. Sci (USA) 93:11288-11294 (1996); Xiao et al.,  
Exp. Neurobiol. 144:113-124 (1997); and Muzyczka, Current Topics in Microbiol. and  
Immunol. 158:97-129 (1992)). Additionally, the packaging cell line can be co-  
30 transfected with a helper plasmid that is an expression plasmid for the AAV proteins  
required *in trans*. This co-transfected packaging cell line can be infected with a helper  
virus so that the packaging cell line produces the recombinant AAV vectors of the present

invention. These AAV vectors can then be used to transduce permissive cells, such as cells in culture, to identify genes and promoters using the methods of the present invention.

The AAV vectors of the present invention can be advantageous for use in the methods of the present invention relative to retrovirus vectors because AAV vectors can be produced at relatively higher titers and can infect relatively quiescent cells compared to retroviral vectors, such as non-lentiviruses.

AAV plasmid vectors are constructed by having various gene or promoter identifying elements (as discussed above for retrovirus vectors) between the two Inverted Terminal Repeats (ITRs) (Rivadeneira et al., Int. J. Oncol. 12:805-810 (1998)). The gene or promoter tagging elements include, but are not limited to reporter genes, splice donor sequences and splice acceptor sequences. Preferably, the reporter gene is adjacent to a splice donor sequence and/or a splice acceptor sequence. The AAV tagging plasmid can be introduced by transfection as is known in the art into packaging cells such as 293 cells. Transient infection of these packaging cell lines with either adenovirus or herpes virus can lead to the generation of AAV particles. The AAV gene-tagging viruses so produced can be used to infect target cells of interest, including relatively quiescent cells, to create a library of cells with the reporter gene integrated into many different genes. The expression profiles of these genes can be monitored by cell sorting the library in the presence and/or absence of a variety of stimuli. Tagged genes or promoters can then be recovered by rapid amplification of cDNA ends (RACE) using known reporter sequences as the anchor for priming and polymerase chain reaction (PCR). Preferred AAV vectors of the present invention are as follows. Like the retrovirus vectors, AAV vectors can have the orientation of these elements reversed.

5' ITR/splice acceptor/beta-lactamase/splice donor/ITR3'  
3' ITR/splice acceptor/beta-lactamase/poly-A/ITR5'  
3' ITR/splice acceptor/IRES/beta-lactamase/poly-A/ITR5'  
3' ITR/splice acceptor/beta-lactamase/poly-A/beta-actin promoter/neo/splice donor/ITR5'  
3' ITR/splice acceptor/beta-lactamase/poly-A/beta-actin promoter/neo/poly-A/ITR5'  
3' ITR/splice acceptor/IRES/beta-lactamase/poly-A/beta-actin promoter/neo/splice donor/ITR5'  
3' ITR/splice acceptor/IRES/beta-lactamase/poly-A/beta-actin promoter/neo/poly-A/ITR5'

3' ITR/splice acceptor/beta-lactamase/poly-A/beta-actin promoter/neo/poly-A/ITR5'  
 3' ITR/splice acceptor/IRES/beta-lactamase/poly-A/beta-actin promoter/neo/poly-A/ITR5'  
 5' ITR/splice acceptor/beta-lactamase/beta-actin promoter/neo/splice donor/ITR3'  
 5' ITR/mutant splice donor/splice acceptor/beta-lactamase/actin promoter/neo/ITR3'  
 5' ITR/splice acceptor/beta-lactamase/beta-actin promoter/neo/ITR3'  
 3' ITR/splice acceptor/IRES/beta-lactamase/poly-A/beta-actin promoter/neo/poly-A/ITR5'  
 3' ITR/splice acceptor/beta-lactamase/poly-A/beta actin promoter/neo/splice donor/ITR5'  
 3' ITR/splice acceptor/IRES/beta-lactamase/poly-A/beta actin promoter/neo/splice donor/ITR5'  
 5' ITR/splice acceptor/reporter gene/eukaryotic promoter/selectable marker/ITR 3'  
 5' ITR/splice acceptor/reporter gene/IRES/eukaryotic promoter/selectable marker/ITR3'  
 3' ITR/splice acceptor/reporter gene/poly-A/eukaryotic promoter/selectable marker/poly-A/ITR5'  
 3' ITR/splice acceptor, reporter gene/IRES/selectable marker/poly-A/ITR5'  
 3' ITR/splice acceptor/reporter gene/poly-A/eukaryotic promoter/reporter gene/splice  
 donor/ITR5'  
 5' ITR/splice acceptor/reporter gene/ITR3'

The present invention also includes methods to amplify genomic regions containing genes or promoters tagged with a reporter gene using a dihydrofolate reductase gene (DHFR). By amplifying the number of reporter genes associated with a gene or promoter, the amount of reporter gene expressed in a cell will increase, which will increase the sensitivity of the detection steps of the methods of the present invention. Vectors containing a DHFR gene preferably will have a wild-type or a methotrexate-resistant variant of a DHFR gene (such as Arg22, Tyr22 or Trp31) associated with a reporter gene in a vector (see, Morris and McIvor, *Biochem. Pharmacol.* 47:1207-1220 (1994)). Initial screening for reporter gene expression in cells can be used to identify clones that express desirable patterns or amounts of reporter gene. These identified clones can then optionally be contacted with increasing concentrations of methotrexate to amplify the genomic regions containing the reporter gene, along with the associated gene or promoter. The result can be a cell line that has more pronounced differential reporter gene expression under different conditions.

For example, a DHFR containing vector can be made by coupling a DHFR gene with a vector of the present invention that includes a reporter gene. Expression of the reporter gene can be under the regulation of the promoter or gene into which the vector will ultimately be inserted within. The linked DHFR gene can also be transcriptionally

regulated from an IRES site, or a promoter provided in the vector. Once a vector has integrated into a gene, a cell or a population of cells can be exposed to sequentially higher concentrations of methotrexate over a period of several days. Surviving cells are expected to have an amplified number of copies of the reporter gene, endogenous promoter, and/or endogenous gene. This aspect of the present invention can be used with any vector of the present invention, such as retroviruses or adeno-associated viruses.

Vectors of the present invention can also be used with liposomes or other vesicles that can transport genetic material into a cell. Appropriate structures are known in the art. The liposomes can include vectors such as plasmids or yeast artificial chromosomes (YACs), which can include genetic material to be introduced into the cell. Plasmids can also be introduced into cells by any known methods, such as electroporation, calcium phosphate, or lipofection. DNA fragments, without a plasmid or viral vector can also be used.

In one aspect of the present invention, vectors are used to introduce reporter genes into cells. When the reporter gene integrates into the genome of a target cell so that the reporter gene is expressed, that event can be detected by detecting the reporter gene. Clones that express the reporter gene under a wide variety of conditions can be used for a variety of purposes, including gene and drug discovery. Chromosomes tagged with beta-lactamase expression constructs can be transferred to desired recipient cells using methods established in the art.

Such vectors can be transformed into appropriate target cells using any appropriate means known in the art, such as lipofection, microbalistics, viral particles, liposomes, electroporation, and the like (see, Sambrook, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press (1989)). Such methods comprise the step of contacting a vector of the present invention with a target cell. Once contacted with the cell, the vector can enter into the target cell where the nucleic acids of the vector can be integrated into the genome of the target cell.

Reporter genes, such as beta-lactamase polynucleotides, can be placed on a variety of plasmids for integration into a genome and to identify genes from a large variety of organisms (Gorman, C.M. et al., Mol. Cell Biol. 2: 1044-1051 (1982); Alam, J. and Cook, J.L., Anal.Biochem. 188: 245-254, (1990)). Standard techniques are used to

introduce these polynucleotides into a cell or whole organism (e.g., as described in Sambrook, J., Fritsch, E.F. and Maniatis, T. Expression of cloned genes in cultured mammalian cells. In: Molecular Cloning, edited by Nolan, C. New York: Cold Spring Harbor Laboratory Press, 1989). Resistance markers can be used to select for successfully transfected cells.

If a beta-lactamase expression construct is selected for integrating a beta-lactamase polynucleotide into a eukaryotic genome, it will usually contain at least a beta-lactamase polynucleotide operably linked to a splice acceptor and optionally a splice donor. Alternatively, the beta-lactamase polynucleotide may be operably linked to any means for integrating a polynucleotide into a genome, preferably for integration into an intron of a gene to produce an in frame translation product. The beta-lactamase expression construct can optionally comprise, depending on the application, an IRES element, a splice donor, a poly A site, translational start site (e.g. a Kozak sequence) an LTR (long terminal repeat) and a selectable marker.

#### *Beta-Lactamase Reporter Genes*

Preferably, beta-lactamase polynucleotides encode a cytosolic form of a protein with beta-lactamase activity. This provides the advantage of trapping the normally secreted beta-lactamase protein within the cell, which enhances signal to noise ratio of the signal associated with beta-lactamase activity. Usually, this is accomplished by removing or disabling the signal sequence normally present for secretion. As used herein, "cytosolic protein with beta-lactamase activity" refers to a protein with beta-lactamase activity that lacks the proper amino acid sequences for secretion from the cell, e.g., the signal sequence. For example, in the polypeptide of **SEQ. ID NO.: 1**, the signal sequence has been replaced with the amino acids Met-Ser. Accordingly, upon expression, beta-lactamase activity remains within the cell. For expression in mammalian cells it is preferable to use beta-lactamase polynucleotides with nucleotide sequences preferred by mammalian cells. In some instances, a secreted form of beta-lactamase can be used with the methods and compositions of the invention. In particular,

genes having sequences that direct secretion can be identified with a beta-lactamase assay. This also permits multiplying based on directed localization of beta-lactamase.

Proteins with beta-lactamase activity can be any known to the art, developed in the future or described herein. This includes, for example, the enzymes represented by  
5 **SEQ. ID. NO.**'s described herein. Nucleic acids encoding proteins with  $\beta$ -lactamase activity can be obtained by methods known in the art, for example, by polymerase chain reaction of cDNA using primers based on the DNA sequence in **SEQ. ID. NO.: 1**. PCR methods are described in, for example, U.S. Pat. No. 4,683,195; Mullis et. al. (1987) *Cold Spring Harbor Symp. Quant. Biol.* 51:263; and Erlich, ed., *PCR Technology*, (Stockton  
10 Press, NY, 1989).

#### *Sequences for Assisting Integration*

The beta-lactamase expression construct typically includes sequences for  
15 integration, especially sequences designed to target or enhance integration into the genome.

The splice site acceptor can be operably linked to the reporter gene (e.g. a beta-lactamase polynucleotide) to facilitate expression upon integration into an intron. Usually, a fusion RNA will be created with the coding region of an adjacent operably  
20 portion of the exon. A splice acceptor sequence is a sequence at the 3' end of an intron where it junctions with an exon. The consensus sequences for a splice acceptor is NTN (TC) (TC) (TC) TTT (TC) (TC)(TC) (TC) (TC) (TC) NCAGgt (see, Shapiro and Senapathy, *Nucleic Acids Research*, 15:7155-7175 (1987)). An example is the splice acceptor sequence from En-2 as described in Gossler, *Nature*, 28 April:463 (1989). The  
25 intronic sequences are represented by upper case and the exonic sequence by lower case font. These sequences represent those that are conserved from viral to primate genomes.

The splice acceptor sequence can be any known in the art (see, for example, Friedrich and Soriano, *Genes & Development*, 5:1513-1523 (1991), Friedrich and Soriano, *Methods in Enzymology*, 225:681-701 (1991), Reddy et al., *Proc. Natl. Acad. Sci. USA* 89:6721-6725 (1992), Wurst et al., *Genetics* 139:889-899 (1995), Hill and  
30 Wurst, *Methods in Enzymology*, 225:664-679 (1993), Shaprio and Senapathy, *Nucleic*

Acids Research, 15:7155-7175 (1987), Gossler et al., Nature 28 April: 463-465 (1989).  
Skarnes et al., Genes and Development, 6:903-918 (1992), and Jarvik et al.,  
BioTechniques 20:896-904 (1990), each of which is incorporated herein by reference).  
The splice donor sequence can be any known in the art (see, for example, Niwa et al., J.  
5 Biochem. 113:343-349 (1993), Yoshida et al., Jarvik et al., BioTechniques 20:896-904  
(1990), and Transgenic Research 4:277-289 (1995), each of which is incorporated herein  
by reference).

The vectors of the present invention can have heterologous splice acceptor  
sequences that can be upstream of the reporter gene. For example, a splice acceptor  
10 having a reduced length, but maintains splice acceptor function can be made using  
methods known in the art. For example, a fragment containing the splice acceptor  
sequence from the engrailed-2 (en-2) gene from Drosophila can be made by reducing the  
size of the 1.8 kb en-2 fragment while maintaining splice acceptor functionality. Such  
splice acceptors having reduced lengths are advantageous in the vectors of the present  
15 invention, such a retroviruses, because smaller vector length can yield higher titers,  
possibly due to increased packaging efficiency or reduced metabolic demand on the  
vector producing cell. Reducing the size of the en-2 splice acceptor is particularly  
desireable because the 3' 100 basepairs of intronic sequence contains the essential  
elements required for splice acceptor activity. Therefore, truncated forms of the en-2  
20 splice acceptor containing only the most 3' 93 basepairs of sequence can be made using  
PCR methods as they are known in the art. The preferred sequence is as follows:

5'-caacctcaagctagcttgggtgcgttggttggtgataagtagctagactccagcaaccagtaacctctgccctttctcctccatgacaaccag-3'

25 The putative splice acceptor branch point is underlined. The size of the splice  
acceptor can be reduced to further reduce or eliminate all sequences upstream of the  
branch point.

As an alternative to a splice donor site, a poly A site may be operably linked to  
the beta-lactamase polynucleotide. Poly-adenylation signals, i.e poly A sites, include  
30 SV40 poly A sites, such as those described in the Invitrogen Catalog 1996 (California).

In some instances, it may be desirable to include in the beta-lactamase expression construct a translational start site. For instance, a translational start site allows for beta-lactamase expression even if the integration occurs in non-coding regions. Usually, such sequences will not reduce the expression of a highly expressed gene. Translational start sites include a "Kozak sequence" and are the preferred sequences for expression in mammalian cells described in Kozak, M., *J. Cell Biol.* 108: 229-241 (1989). The nucleotide sequence for a cytosolic protein with  $\beta$ -lactamase activity in **SEQ. ID. NO.: 3** contains a Kozak sequences for the nucleotides -9 to 4 (GGTACCACCATGA).

It is also preferable, when using mammalian cells, to include an IRES ("internal ribosome entry binding site") element in the beta-lactamase or reporter gene expression construct. Typically, an IRES element will improve the yield of expressing clones. One caveat of integration vectors is that only one in three insertions into an intron will be in frame and produce a functional reporter protein. This limitation can be reduced by cloning an IRES sequence between the splice acceptor site and the reporter gene (e.g., a  $\beta$ -lactamase polynucleotide). This eliminates reading frame restrictions and possible functional inactivation of the reporter protein by fusion to an endogenous protein. IRES elements include those from picornaviruses, picorna-related viruses, and hepatitis A and C. Preferably, the IRES element is from a poliovirus. Specific IRES elements can be found, for instance, in WO9611211 by Das and Coward published 4/16/96, EP 585983 by Zurr published 3/7/96, WO9601324 by Berlioz published 1/18/96 and WO9424301 by Smith published October 27, 1994, all of which are herein incorporated by reference. To improve selection of beta-lactamase polynucleotide into a genome, a selectable marker can be used in the beta-lactamase expression construct. Selectable markers for mammalian cells are known in the art, and include for example, thymidine kinase, dihydrofolate reductase (together with methotrexate as a DHFR amplifier), aminoglycoside phosphotransferase, hygromycin B phosphotransferase, asparagine synthetase, adenosine deaminase, metallothionien, and antibiotic resistant genes such as genes for neomycin, hygromycin and puromycin resistance. Selectable markers for non-mammalian cells are known in the art and include genes providing resistance to antibiotics, such as kanamycin, tetracycline, and ampicillin.



The invention can be readily practiced with genomes having intron/exon structures. Such genomes include those of mammals (e.g., human, rabbit, mouse, rat, monkey, pig and cow), vertebrates, insects and yeast. Intron-targeted vectors are more commonly used in mammalian cells as introns, or intervening sequences, are considerably larger than exons, or mRNA coding regions in mammals. Intron targeting can be achieved by cloning a splice acceptor or 3' intronic sequences upstream of a  $\beta$ -lactamase polynucleotide gene followed by a polyadenylation signal or 5' intronic splice donor site. When the vector inserts into an intron, the reporter gene (e.g.,  $\beta$ -lactamase) is expressed under the same control as the gene into which it has inserted.

The invention can also be practiced with genomes having reduced numbers of, or lacking, intron/exon structures. For lower eukaryotes, which have simple genomic organization, i.e. containing few and small introns, exon-targeted vectors can be used. Such vectors include  $\beta$ -lactamase polynucleotides operably linked to a poly-adenylation sequence and optionally to an IRES element. Lower eukaryotes include yeast, and fungi and pathogenic eukaryotes (e.g. parasites and microorganisms). For genomes lacking intron/exon structures restriction enzyme integration, transposon induced integration or selection integration can be used for genomic integration. Such methods include those described by Kuspa and Loomis, PNAS 89: 8803-8807 (1992) and Derbyshire, K.M., Gene Nov. 7: 143-144 (1995). Prokaryotes can be used with the invention if integration can occur in such genomes. Retroviral vectors can also be used to integrate  $\beta$ -lactamase polynucleotides into a genome (e.g., eukaryotic), such as those methods and composition described in U.S. Patent Number 5,364,783.

Typically, integration will occur in the regions of the genome that are accessible to the integration vector. Such regions are usually active portions of the genome where there is increased genome regulatory activity, e.g. increased polymerase activity or a change in DNA binding by proteins that regulate transcription of the genome. Many embodiments of the invention described herein can result in random integration, especially in actively transcribed regions.

### *Integration into Active Portions of the Genome*

Integration, however, can be directed to regions of the genome active during specific types of genome activity. For instance, integration at sites in the genome that are active during specific phases of the cell cycle can be promoted by synchronizing the cells in a desired phase of the cell cycle. Such cell cycle methods include those known in the art, such as serum deprivation or alpha factors (for yeast). Integration may also be directed to regions of the genome active during cell regulation by a chemical, such as an antagonist or agonist for a receptor or some other chemical that increase or decreases or otherwise modulates genome activity. By adding the chemical of interest, genome activity can be increased, often in specific regions to promote integration of an integration vector (e.g. as a reporter gene construct), including those of the invention, into such regions of the genome.

For instance, a nuclear receptor activator (general or specific) could be applied to activate the cells prior or during integration in order to promote integration of reporter genes at sites in the genome that become more active during nuclear receptor activation. Such cells could then be screened with the same or different nuclear receptor activator to identify which clones, and which portions of the genome are active during nuclear receptor activation. Any agonists, antagonists and modulators of the receptors described herein can be used in such a manner, as well as any other chemicals that increase or decrease genome activity.

### *Cells for Integration into the Genome*

The cells used in the invention will typically correspond to the genome of interest. For example, if regions of the human genome are desired to be identified, then human cells containing a proper genetic complement will generally be used. Libraries, however, could be biased by using cells that contain extra-copies of certain chromosomes or other portions of the genome. Cells that do not correspond to the genome of interest can also be used if the genome of interest or significant portions of the genome of interest can be replicated in the cells, such as making a human-mouse hybrid.

Additionally, by the appropriate choice of cells and expressed proteins, identification and screening assays can be constructed that detect active portions of the genome associated with a biological process that requires, in whole or part, the presence of a particular protein (protein of interest). Cells can be selected depending on the type of proteins that are expressed (homologously or heterologously) or from the type of tissue from which the cell line or explant was originally generated. If the identification of portions of the genome activated by a particular type of protein is desired, then the cell used should express that protein.

The cells can express the protein homologously, i.e. expression of the desired protein normally or naturally occurs in the cells. Alternatively, the cells can be directed to express a protein heterologously, i.e. expression of the desired protein which does not normally or naturally occur in the cells. Such heterologous expression can be directed by “turning on” the gene in the cell encoding the desired protein or by transfecting the cell with a polynucleotide encoding the desired protein (either by constitutive expression or inducible expression). Inducible expression is preferred if it is thought that the expressed protein of interest may be toxic to the cells.

Many cells can be used with the invention. Such cells include, but are not limited to adult, fetal, or embryonic cells. These cells can be derived from the mesoderm, ectoderm, or endoderm and can be stem cells, such as embryonic or adult stem cells, or adult precursor cells. The cells can be of any lineage, such as vascular, neural, cardiac, fibroblasts, lymphocytes, hepatocytes, cardiac, hematopoietic, pancreatic, epidermal, myoblasts, or myocytes. Other cells include baby hamster kidney (BHK) cells (ATCC No. CCL10), mouse L cells (ATCC No. CCL1.3), Jurkats (ATCC No. TIB 152) and 153 DG44 cells (see, Chasin (1986) Cell. Molec. Genet. 12: 555) human embryonic kidney (HEK) cells (ATCC No. CRL1573), Chinese hamster ovary (CHO) cells (ATCC Nos. CRL9618, CCL61, CRL9096), PC12 cells (ATCC No. CRL17.21) and COS-7 cells (ATCC No. CRL1651). Preferred cells include Jurkat cells, CHO cells, neuroblastoma cells, P19 cells, F11 cells, NT-2 cells, and HEK 293 cells, such as those described in U.S. Patent No. 5,024,939 and by Stillman et al. Mol. Cell. Biol. 5: 2051-2060 (1985). Preferred cells for heterologous protein expression are those that can be readily and efficiently transfected.

Cells used in the present invention can be from continuous cell lines or primary cell lines obtained from, for example, mammalian tissues, organs, or fluids. Tissue sections as well as disperse cells can be used in the present invention. Cells can also be obtained from transgenic animals that have been engineered to express a reporter gene.

5 Cells obtained from transgenic or non-transgenic animals are preferred for cells that are difficult to culture *in vitro*, such as neural and hepatic cells. Primary cell lines can be made continuous using known methods, such as fusing primary cells with a continuous cell line or expressing transforming proteins. Cells of the invention can be stored or used with methods of the invention as isolated, clonal populations in plates such as those

10 described in commonly owned United States Patent Applications having Attorney Docket Nos: 08366/010001, entitled "Low background multi-well plates and platforms for spectroscopic measurements" (Coassin et al., filed June 2, 1997); and 08366/009001, entitled "Low background multi-well plates with greater than 864 wells for spectroscopic measurements" (Coassin et al., filed June 2, 1997); each of which is incorporated herein

15 by reference plates. Preferably, cells are stored or used in plates with 96, 384, 1536 or 3456 wells per plate. A single cell or a plurality of cells can be placed in such wells. Such isolated clonal populations will typically have 1,000, 10,000, or 100,000 or more such populations representative of substantially equivalent numbers of independent integrations sites. Such panels can be used in profiling, pathway identification,

20 modulator identification, modulator characterization, and other methods of the invention.

Another aspect of the present invention is a cell that comprises a vector of the present invention. The cells of the present invention can be made by transfecting or infecting a target host cell with a vector of the present invention. Target cells can be any eukaryotic cell, preferably from an organism such as a plant, insect, or mammalian cells

25 such as human cells. The eukaryotic cell can also be any unicellular eukaryotic cell, such as a yeast cell or other unicellular organism. When transfected, the target cells can be in a living or non-living organism, in an isolated tissue, organ, or fluid from an organism, or cells isolated from an organism. Target cells can be obtained from any tissue, fluid, or organ of a plant or animal and can be primary or continuous cell lines. Continuous cell

30 lines can be made using methods known in the art, such as fusing primary cells with a

continuous cell line. Animals, such as knock-out mice, can also be made from mice having appropriate vectors described herein.

Prior to or after transfection with a trapping vector of the present invention, cells can be transfected with an exogenous gene capable of expressing an exogenous protein, such as a receptor (e.g., GPCR) or gene associated with the pathology of an etiological agent, such as a virus, bacteria, or parasite. Cells that express such exogenous proteins can then be transfected with a trapping vector to form a library of clones that can be screened using the present invention. The invention can also include animals with  $\beta$ -lactamase expression or reporter gene constructs integrated into the genome of interest.

Many of the cells of the present invention can report modulation of biological processes by a variety of additional reporter genes or chemicals or combinations thereof. For example, beta-lactamase, an enzyme, can convert non-chromogenic substrates to chromogenic products or alter the chromogenic or fluorescent properties of a substrate such as CCF2. Furthermore, fluorescent reporters, such as fluorescent proteins, such as green fluorescent protein (GFP) molecules, can be used as reporters. Some mutant GFP molecules have different fluorescent properties as compared to wild-type GFP. These GFPs can be used as reporters and can be used singly or in combination with the present invention. For example, cells can have multiple reporters that can be differentiated to report different biological processes, or different steps within a biological process, such as steps in a signal transduction pathway.

### *Targets*

Proteins of interest that can be expressed in the cells of the invention include: hormone receptors (e.g. mineralcorticosteroid, glucocorticoid, and thyroid hormone receptors); intramembrane proteins (e.g. TM-1 and TM-7) intracellular receptors (e.g., orphans, retinoids, vitamin D3 and vitamin A receptors); signaling molecules (e.g., kinases, transcription factors, or molecules such signal transducers and activators of transcription) (*Science* Vol. 264, 1994, p.1415-1421; *Mol. Cell Biol.*, Vol. 16, 1996, p.369-375); receptors of the cytokine superfamily (e.g. erythropoietin, growth hormone, interferons, and interleukins (other than IL-8) and colony-stimulating factors); G-protein

coupled receptors, see US patent 5,436,128 (e.g., for hormones, calcitonin, epinephrine, gastrin, and pancrein or autocrine mediators, such as stomatostatin or prostaglandins) and neurotransmitter receptors (norepinephrine, dopamine, serotonin or acetylcholine); tyrosine kinase receptors (such as insulin growth factor, nerve growth factor (US patent 5,436,128)). Examples of the use of such proteins is further described herein.

Any target, such as an intracellular or extracellular receptor involved in a signal transduction pathway, such as the leptin or GPCR pathways, can be used with the present invention. Furthermore, the genes activated or repressed by a target can be isolated, identified, and modulators of that gene identified using the present invention. For example, the present invention can identify a G-protein coupled receptor (GPCR) pathway, determine its function, isolated the genes modulated by the GPCR, and identify modulators of such GPCR modulated proteins.

As an introduction to GPCR cell biology, the activation of  $G\alpha_{15}$  or  $G\alpha_{16}$  can, through a G-protein signaling pathway, activate  $PLC\beta$ , which in turn increases intracellular calcium levels. An increase in calcium levels can lead to modulation of a “calcium-responsive” promoter that is part of a signal transduction detection system, i.e., a promoter that is activated (e.g., a NFAT promoter AP-1) or inhibited by a change in calcium levels. One example of an NFAT DNA binding site is described in Shaw, et al. Science 291:202-205 (1988). Likewise, a promoter that is responsive to changes in protein kinase C levels (e.g., a “protein kinase C-responsive promoter”) can be modulated by an active  $G\alpha$  protein through G-protein signaling pathway. Selected cells described herein can also include a G-protein coupled receptor. Genes encoding numerous GPCRs have been cloned (Simon et al., Science 252:802-808 (1991)), and conventional molecular biology techniques can be used to express a GPCR on the surface of a cell of the invention. Preferably, the sum responsive promoter can allow for only a relatively short lag (e.g., less than 90 minutes) between engagement of the GPCR and transcriptional activation. A preferred responsive promoter includes the nuclear factor of activated T-cell promoter (Flanagan et al., Nature 352:803-807 (1991)). Polynucleotides identified by methods of the invention can be used as response elements that are sensitive to intracellular signals (signal-response elements). Signal response elements can be used in the assays described herein, such as identification of useful chemicals. Such signal

response elements may sensitive intracellular signals that include voltage, pH, and intracellular levels of  $\text{Ca}^{++}$ , ATP, ADP, cAMP, GTD, GDP,  $\text{K}^+$ ,  $\text{Na}^+$ ,  $\text{Zn}^{++}$ , oxygen, metabolites and  $\text{IP}_3$ .

In one aspect of the present invention, cells can be transformed to express an exogenous receptor, such as GPCR. Such a transduced cell line can than be further transduced with a trapping vector to make a library of clones that can be used to identify cells that report modulation of the exogenous receptor. Preferably, the host cell line would not appreciably express the exogenous receptor.

Based on the unique structure of GPCRs, which have seven hydrophobic, presumably trans-membrane, domains (see, Watson and Arkinstall, The G-Protein Linked Receptor Facts Book, Academic Press, New York (1994)) orphan GPCRs (GPCRs having no known function) can be identified by searching sequence databases, such as those provided by the National Library of Medicine (Bethesda, MD), for similar motifs and homologies. This same strategy can, of course, be used for any target, especially when a paradigm sequence or motif has been determined.

#### *Drug Discovery for Viruses and Other Pathogens*

The function of genes from viruses or other pathogens that effect the expression of genes in cells, such as mammalian cells, can be determined using the present invention. Furthermore, chemicals that modulate these genes can be identified using the methods of the present invention. For example, many transforming viruses, after infecting a cell, have the effect of up-regulating genes involved in cell proliferation, which allows the virus-infected cells to produce additional viruses, which can infect additional cells.

These transforming viruses can act by stimulating a receptor from the target cell. One example of the mechanism is the Friend Erythroleukemia virus. This virus uses the erythropoietin receptor for entry into the cells. When the virus is bound to the receptor, a pathway is activated that causes an over-proliferation of red blood cells. If the activation of the erythropoietin receptor is inhibited, a decrease in the accumulation of red blood cells would result which can prevent or reduce the severity of the leukemia. The development of an assay that reports the activation of mammalian target genes allows the

identification of modulators of other viral or pathogenic dependent pathways. These modulators can be used as therapeutic agents.

A general procedure for establishing this assay uses the virus or an isolated viral protein as the stimulus for modulating a pathway. First, a gene-trapping library is made using a cell line that can be infected by the virus or activated by the viral protein. The virus is added to these cells, and clones are isolated that responded specifically to the viral infection by the expression of a reporter gene.

As an example, the GP-120 portion of HIV protein is known to have mitogenic effect on cells exposed to GP-120, which indicates that downstream signaling pathways are being activated that can be associated with the cytotoxicity of the virus and allow its proliferation. Cell clones can be isolated that are induced by this activation which can be used to screen for modulators of this cytotoxic or proliferative effect. Other viral proteins, such as NEF from HIV, can be used. Chemicals that inhibit this effect can have useful therapeutic value to treat viral infection or toxicity.

This approach can be applied to any cellular pathogen that has an effect on a target cells, such as cytotoxicity, cell proliferation, inflammation or other responses. Other etiological targets include other viruses, such as retroviruses, adenovirus, papillomavirus, herpesviruses, cytomegalovirus, adeno associated viruses, hepatitis viruses, and any other virus. In addition to viruses, any other pathogen, such as parasites, bacteria, and viroids, can be used in the present invention. Particular viral targets include, but are not limited to, NEF, Hepatitis X protein, and other viral proteins, such as those that can be encoded or carried by a virus. In addition, two or more viral components can be added to identify coviral pathogenesis components. This is a particularly valuable tool for identifying pathways modulated by two or more viruses concurrently, or over time as in slow activating viral conditions. For example, cotransfection with HIV and CMV may be used. Viral targets or components do not include oncogenes or proto-oncogenes found in uninfected genomes, and gene products thereof.



*Screening Test Chemicals Using Portions Of The Genome*

Cells comprising beta-lactamase polynucleotides integrated in the genome can be contacted with test chemicals or modulators of a biological process and screened for activity. Usually, the test chemical being screened will have at least one defined target, usually a protein. The test chemical is normally applied to the cells to achieve a final predetermined concentration in the medium bathing the cells. Typically, screens are conducted at concentrations 100 microM or less, preferably 10 microM or less and preferably 1 microM or less for confirmatory screens. As described more fully herein, cells can be subjected to multiple rounds of screening and selection using the same chemical in each round to insure the identification of clones with the desired response to a chemical or with different chemicals to characterize which chemicals produce a response (either an increase or decrease beta-lactamase activity) in the cells. Such methods can be applied to any chemical that alters the function of any the proteins mentioned herein or known in the art.

Chemicals and physiological processes without a defined target, however, can also be used and screened with the cells of the invention. For example, once a clone is identified as containing an active genomic polynucleotide that is activated by a particular cellular signal (including extracellular signals), for instance by a neurotransmitter, that same clone can be screened with chemicals lacking a defined target to determine if activation by the neurotransmitter is blocked or enhanced by the chemical. This is a particularly useful method for finding therapeutic targets downstream of receptor activation (in this case a neurotransmitter). Such methods can be applied to any chemical that alters the function of any the proteins mentioned herein or known in the art. This type of "targetless" assay is particular useful as a screening tool for the medial conditions and pathways described herein.

The methods and compositions described herein offer a number of advantages over the prior art. For instance, screening of mammalian based gene integration libraries is limited by the use of existing reporter systems. Many enzymatic reporter genes, such as secreted-alkaline phosphatase, and luciferase, cannot be used to assay single living cells (including FACS) because the assay requires cell lysis to determine reporter gene

activity. Alternatively, beta-galactosidase can detect expression in single cells but substrate loading requires permeabilization of cells, which can cause deleterious effects on normal cell functions. Additionally, the properties of fluorescent beta-galactosidase substrates, such as fluoroscein di-beta-D-galactopyranside, and products make it very difficult to screen large libraries for both expressing and non-expressing cells because the substrate and product is not well retained or permits ratiometric analysis to determine the amount uncleaved substrate. Green fluorescent protein (GFP), a non-enzymatic reporter, could be used to detect expression in single living cells but has limited sensitivity. GFP expression level would have to be at least 100,000 molecules per cell to be detectable in a screening format and small changes in, or low levels of, gene expression could not be measured. Furthermore GFP is relatively stable and would not be suitable for measuring down-regulation of genes. Other advantages of the invention are described herein or readily recognized by one skilled in the art upon reviewing this disclosure.

#### **Methods for Rapidly Identifying Modulators of Genomic Polynucleotides**

The invention provides for a method of identifying proteins or chemicals that directly or indirectly modulate a genomic polynucleotide. Generally, the method comprises inserting a beta-lactamase expression construct into an eukaryotic genome, usually non-yeast, contained in at least one living cell, contacting the cell with a predetermined concentration of a modulator, and detecting beta-lactamase activity in the cell. Preferably, cleavage of a membrane permeant beta-lactamase substrate is measured and the membrane permeant beta-lactamase substrate is transformed in the cell into a trapped substrate. Preferably, the beta-lactamase expression construct comprises a  $\beta$ -lactamase polynucleotide, a splice donor, a splice acceptor and an IRES element. The method can also include determining the coding nucleic acid sequence of a polynucleotide operably linked to the  $\beta$ -lactamase expression construct using techniques known in the art, such as RACE.

## *Modulator Identification*

Modulators described herein can be used in this system to test for an increase or decrease in beta-lactamase activity in successfully integrated clones. Such cells can optionally include specific proteins of interest as discussed herein. For example, the cell can include a protein or receptor that is known to bind the modulator (e.g., a nuclear receptor or receptor having a transmembrane domain heterologously or homologously expressed by the cell). A second modulator can be added either simultaneously or sequentially to the cell or cells and beta-lactamase activity can be measured before, during or after such additions. Cells can be separated on the basis of their response to the modulator (e.g. responsive or non-responsive) and can be characterized with a number of different modulators to create a profile of cell activation or inhibition.

Beta-lactamase activity will often be measured in relation to a reference sample, often a control. For example, beta-lactamase activity is measured in the presence of the modulator and compared to the beta-lactamase activity in the absence of the modulator or possibly a second modulator. Alternatively, beta-lactamase activity is measured from a cell expressing a protein of interest and to a cell not expressing the protein of interest (usually the same cell type). For instance, a modulator may be known to bind to a receptor expressed by the cell and the beta-lactamase activity in the cell is increased in the presence of the modulator compared to the beta-lactamase activity detected from a corresponding cell in the presence of the modulator, wherein the corresponding cell does not express the receptor.

## *Pathway Identification and Modulators*

When a reporter gene of the invention integrates into the genome of a host cell such that the reporter gene is expressed under a variety of circumstances, these clones can be used for drug discovery and functional genomics. These clones report the modulation of the reporter gene in response to a variety of stimuli, such as hormones and other physiological signals. These stimuli can be involved in a variety of known or unknown pathways that are modulated by known or unknown modulators or targets. Thus, these

clones can be used as a tool to discover chemicals that modulate a particular pathway or to determine a cellular pathway.

These pathways are quite varied, and fall into general classes, which have specific species, which can be modulated by known or unknown modulators or agonists or antagonists thereof. By way of example, **Table 1** illustrates various pathways, species of these classes, and known modulators of these species. The invention can be used to identify regions of the genome that are modulated by such pathways, or physiological event

**TABLE 1**

**Pathways and modulators**

Pathway/Physiological Event		
Genus	Species	Known Modulator
Nuclear receptors	Estrogen receptor	Estrogen
Cytokines	IL-2 receptor	IL-2
GPCRs	Vasopressin receptor	Vasopressin
Transcription factors	Fos or Jun	NFAT
Kinase dependent	Protein kinase C	PMA
Phosphatase dependent	Calcineurine	Cyclosporin A
Protease dependent	Metalloprotease	TIMPs
Chemokine	CCR1	RANTES
Ion channels	Calcium channels	Many known blockers
Second messenger dependent	Cyclic AMP	CAMP inhibitor protein
Cell differentiation	Hematopoietic development	EPO
Cell growth	IL-2 receptor	IL-2
Cell cycle dependent	CDK	P21
Apoptosis	Fas	P53

In one embodiment, the invention provides for a genomic assay system to identify downstream transcriptional targets for signaling pathways. This method requires the target of interest to activate gene expression upon addition of chemical or expression of the target protein. A cell line that is the most similar to the tissue type where the target functions is preferred for generating a library of clones with different integration sites with  $\beta$ -lactamase polynucleotides or other reporter genes. This cell line may be known to elicit a cellular response, such as differentiation upon addition of a particular modulator.

If this type of cell line is available, it is preferred for screening, as it represents the native context of the target. If a cell line is not available that homologously expresses the target; a cell line can be generated by heterologously expressing the target in the most relevant cell line. For instance, if the target is normally expressed in the lymphoid cells, then a lymphoid cell line would be used generate the library.

The library of clones, as described further herein, can be separated into two pools by FACS using the FRET system described herein: an expressing pool (e.g. blue cells) and a non-expressing pool (e.g. green cells). These two pools can then be treated with a modulator followed by FACS to isolate induced clones (e.g. green to blue) or repressed clones (e.g. blue to green). Additional rounds of stimulation followed by FACS can be performed to verify initial results. The specificity of activation can be tested by adding additional chemicals that would not activate the defined target. This would allow the identification of clones that have  $\beta$ -lactamase polynucleotides integrated into genes activated by a variety of cellular signals.

Once a pool of cells with the desired characteristics are isolated they can be expanded and their corresponding genes cloned and characterized. Targets that could be used in this assay system include receptors, kinases, protein/protein interactions or transcription factors and other proteins of interest discussed herein.

Another aspect of the present invention is a library of cells made by a method of the present invention. The library of cells can be a pool of cells, such as before or after FACS sorting. Alternatively, a library of cells can be separate individual clones, or clonal population, that are kept separate. These individual clones or clonal populations can be present in a two dimensional array, such as in a multi-well platform, such as a microtiter plate, having a different clone, clonal population, or population of cells in each well. Alternatively, the two-dimensional matrix can be a gel, such as an agarose or alginate-based gel. Libraries of populations of cells preferably have between about 1,000 members and about 10,000,000 members, more preferably between about 100,000 members and about 8,000,000 members, and most preferably between about 1,000,000 members and about 5,000,000 members. Libraries of individual clones or clonal populations preferably have between about 10 members and about 10,000 members, more preferably between about 50 members and about 5,000 members, and most

preferably between about 100 and about 1,000 members.

In another embodiment, the invention provides for a method of identifying developmentally or tissue specific expressed genes.  $\beta$ -lactamase polynucleotide can be inserted, usually randomly, into any precursor cell such as an embryonic or hematopoietic stem cell to create a library of clones. Constitutively expressing clones can be collected by sorting for blue cells and non-expressing cells collected by sorting for green cells using the FRET system described herein. The library of clones can then be stimulated or allowed to differentiate, and induced or repressed clones isolated. Cell surface markers in conjunction with fluorescent tagged antibodies or other detector molecules could be used to monitor the expression of reference genes simultaneously. Additionally, by stimulation and sorting stem cells at various developmental stages, it is possible rapidly identify genes responsible for maturation and differentiation of particular tissues.

Additionally, clones that have a beta-lactamase polynucleotide integrated, either randomly or by homologous recombination, into developmentally expressed genes can be used with FACS to isolate specific cell populations for further study, such as screening. Such methods can be used for identifying cell populations that have stem cells properties, as well as providing an intracellular reporter that allows isolation and screening of such a population of cells.

The present invention can yield screening cell lines for a variety of targets whose downstream signaling elements are already known or postulated. These screening cell lines can be used to either screen for modulators of transfected targets or as readouts for expression cloning or functional analysis of uncharacterized targets. Screening cell lines can be made for any pathway or any modulator, such those described in **Table 1**.

In the case of ion channels, cell lines are generated in which beta-lactamase expression is used to detect a voltage change. This is possible because intracellular signaling is sensitive to membrane potential and will modulate the expression of a subset of genes. In one example, a library of neuronal cells prepared following the general methods set forth in Examples 1 to 13, such as a dorsal root neuroblastoma cells, be screened for a response to a depolarization by incubating cells in high potassium (high  $K^+$ ) medium. Depending on the particular characteristics of the cell library and the method used, clones with a transcriptional response to a depolarizing treatment are

identified by sorting for cells which changed from either green to blue or blue to green after depolarization. These clones are designated as voltage-sensitive clones and can be used as screening cell lines to identify chemicals that modulate ion channels (either endogenously expressed or transfected) which cause a voltage change upon either activation or inhibition (e.g.  $K^+$  or  $Na^+$  channels). These cells are also useful for expression cloning of ion channels. For example, a voltage-sensitive clone could be transfected with a cDNA library. Those cells transfected with functional channels that shift the membrane potential are detected via beta-lactamase and the cDNA gene products are analyzed for activity as ion channels.

Furthermore, a gene encoding a known ion channel can be transfected into the voltage sensitive cell line and then used as a screen for channel modulators. For example, expression or pharmacological activation of a  $Na^+$  channel can cause a depolarization that can be reported by the cell line. This cell line can be used to screen for agonists or antagonists, depending on the experimental protocol of ion channel modulators. In a variation of this approach, a genomic library from a cell line lacking  $K^+$  channels, such as L929 cells, can be directly transfected with a  $K^+$  channel gene. The expression of the  $K^+$  channel causes a voltage shift, such as a hyperpolarization, causing a change in expression of certain voltage-sensitive genes. The clones expressing these genes can be used to screen for regulators of the ion channel.

#### *Orphan protein signaling pathway identification and orphan protein modulators*

In another embodiment, the invention provides for a method of identifying modulators of orphan proteins or genomic polynucleotides that are directly or indirectly modulated by an orphan protein. Human disease genes are often identified and found to show little or no sequence homology to functionally characterized genes. Such genes are often of unknown function and thus encode for an "orphan protein." Usually such orphan proteins share less than 25% amino acid sequence homology with other known proteins or are not considered part of a gene family. With such molecules there is usually no therapeutic starting point. By using libraries of the herein described clones, one can extract functional information about these novel genes.

Orphan proteins can be expressed, preferably overexpressed, in living mammalian cells. By inducing over expression of the orphan gene and monitoring the effect on specific clones one may identify genes that are transcriptionally regulated by the orphan protein. By identifying genes whose expression is influenced by the novel disease gene or other orphan protein one may predict the physiological bases of the disease or function of the orphan molecule. Insights gained using this method can lead to identification of a valid therapeutic target for disease intervention.

*Modulator Identification using Genomic Polynucleotides Activated by Cellular Signals*

In another embodiment, the invention provides for a method of screening a defined target or modulator using genomic polynucleotides identified with the methods described herein. The gene identification methods described herein can also be used in conjunction with a screening system for any target that functions (either naturally or artificially) through transcriptional regulation.

In many instances a receptor and its ligand are known but not the downstream biological processes required for signaling. For example, a cytokine receptor and cytokine may be known but the downstream signaling mechanism is not. A library of clones generated from a cell line that expresses the cytokine receptor can be screened to identify clones showing changes in gene expression when stimulated by the cytokine. The induced genes could be characterized to describe the signaling pathway. Using the methods of the invention, gene characterization is not required for screen development, as identification of a cell clone that specifically responds to the cytokine constitutes a usable secondary screen. Therefore, clones that show activation or deactivation upon the addition of the cytokine can be expanded and used to screen for agonists or antagonists of cytokine receptor. The advantage of this type of screening is that it does not require an initial understanding of the signaling pathway and is therefore uniquely capable of identifying leads for novel pathways.

In another embodiment, the invention provides for a method of functionally characterizing a target using a panel of clones having active genomic polynucleotides as identified herein. As large numbers of specifically responding cell lines containing active



genomic polynucleotides identified with a particular biological process or modulator are generated. panels containing specific clones can be used for functional analysis of other potential cellular modulators. These panels of responding cell lines can be used to rapidly profile potential transcriptional regulators. Such panels, as well as containing clones with identified active genomic polynucleotides, which were generated by the invention panels, can include clones generated by more traditional methods. Clones can be generated that contain both the identified active genomic polynucleotide with a  $\beta$ -lactamase polynucleotide and specific response elements, such as SRE, CRE, NFAT, TRE, IRE, or reporters under the control of specific promoters. These panels would therefore allow the rapid analysis of potential effectors and their mechanisms of cellular activation. A second reporter (e.g.  $\beta$ -galactosidase gene can also be used with this method, as well as the other method described herein.

In another embodiment, the invention provides for a method of test chemical profiling using a clone or panel of clones having identified active polynucleotides. Test chemical characterization is similar to target characterization except that the cellular target(s) do not have to be known. This method will therefore allow the analysis of test chemical (e.g. lead drugs) effects on cellular function by defining genes effected by the drug or drug lead.

Such a method can find application in the area of drug discovery and secondary affects (e.g. cytotoxic affect) of drugs. The potential drug would be added to a library of genomic clones and clones that either were induced or repressed would be isolated, or identified. This method is analogous to target characterization except that the secondary drug target is unknown. As well as providing a screen for the secondary effects, the assay provides information on the mechanism of toxicity.

#### *Methods Related to FACS and Identifying Active Genomic Polynucleotides*

The invention provides for a method of identifying active genomic polynucleotides using clones having integrated beta-lactamase polynucleotides and FACS. Beta-lactamase integration libraries can be used in a high-throughput screening format, such as FACS, to detect transcriptional regulation. The compatibility of beta-

lactamase assays with FACS enables a systematic method for defining patterns of transcriptional regulation mediated by a range of factors. This approach has not been feasible or practical using existing reporter systems. This new method will allow rapid identification of genes responding to a variety of signals, including tissue specific expression and during pattern formation.

For example, after integration of a beta-lactamase polynucleotide, expressing and non-expressing cells can be separated by FACS. These two cell populations can be treated with potential modulators and changes in gene expression can be monitored using ratio-metric fluorescent readout. Pools of clones will be isolated that show either up- or down-regulation of reporter gene expression. Target genes from responding clones can then be identified. In addition, by being able to separate expressing and non-expressing cells at different time points after modulator addition, genes that are differentially regulated over time can be identified. This approach therefore enables the elucidation of transcription cascades mediated by cellular signaling. Specifically, it will provide a means to identify downstream genes which are transcriptionally regulated by a variety of molecules including, nuclear receptors, cytokine receptors or transcription factors.

Applications of this technology are nearly unlimited in the areas of gene discovery and functional analysis. Libraries of cell lines from various tissue types could be generated and used to identify genes with specific expression patterns or regulation mechanisms. These libraries of clones would represent millions of integration sites saturating the genome and can permit the identification of any expressed gene based on its transcriptional regulation. The features of the  $\beta$ -lactamase reporter system, in part, allow its use for this genomic integration assay in a high-throughput format

There are a variety of other approaches that may be used with the invention, including approaches similar to those proposed for  $\beta$ -lactamase. Examples would include antibody epitopes presented on the cell surface with fluorescent antibodies to detect positive cells. Gel matrixes could also be used which retain secreted reporters and allow detection of positive cells. These approaches would, however, be limited in sensitivity and would not be ratiometric in their detection. They would therefore allow for only the sorting of positive cells based on fluorescent intensity.

Once active genomic polynucleotides have been identified, they can be sequenced using various methods, including RACE (rapid amplification of cDNA ends). RACE is a procedure for the identification of unknown mRNA sequences that flank known mRNA sequences. Both 5' and 3' ends can be identified depending on the RACE conditions.

5' RACE is done by first preparing RNA from a cell line or tissue of interest. This total or polyA RNA is then used as a template for a reverse transcription reactions which can either be random primed or primed with a gene-specific primer. A poly nucleotide linker of known sequence is then attached to the 3' end of the newly transcribed cDNA by terminal transferase or RNA ligase. This cDNA is then used as the template for PCR using one primer within the reporter gene and the other primer corresponding to sequence which had been linked to the 3' end of the first stand cDNA. The present invention is particularly well suited for such techniques and does not require construction of additional clones or constructs once the genomic polynucleotide has been identified.

The splice donor site can be operably linked to the reporting gene (e.g.  $\beta$ -lactamase polynucleotide) or a selectable marker to facilitate integration in an intron to promote expression stability of the mRNA transcript by using an endogenous downstream poly-A sequence. Usually, a fusion RNA is created with the coding region or untranslated on the 3' end of the  $\beta$ -lactamase polynucleotide or selectable marker. This is preferred when it is desired to sequence the coding region of the identified gene.

A splice donor is a sequence at the 5' end of an intron where it junctions with an exon. The consensus sequence for a splice donor sequence is naggGT(A or G)AGT (see, Shapiro et al., Nucl. Acids. Res. 17:7155 (1987)). Other appropriate splice donor sequences are gagGTAAGTA and cagGTGAGTTCGCAT (the complete sequence from the beta-actin gene is reported by Cover, Nucleic Acids Res. 11:1759-1771 (1983) (see, positions 1687 to 2114)). The intronic sequences are represented by upper case and the exonic sequence by lower case font. These sequences represent those that are conserved from viral to primate genomes. This splice donor allows identification of the target gene using 3' RACE. The 3' RACE method (Frohman et al., Proc. Natl. Acad. Sci USA 85: 8993-9002(1988)) is useful for finding the 3' end of a nucleic acid sequence when the sequence upstream of the 3' end of a nucleic acid sequence is known. As used in the present invention, 3' RACE allows the rapid identification of endogenous genes isolated

by the methods of present invention. In practice, RNA (either total RNA or mRNA) can be isolated from a clone or pool of clones identified by a method of the present invention.

The RNA is reverse transcribed using an oligo-dT primer using methods known in the art. The first strand of DNA obtained by reverse transcription can be used as a

5 template for PCR reactions using an oligo-dA primer and a primer that corresponds to at least a portion of a vector of the present invention. The choice of a useful second primer can be made based on the state of the art of PCR methods (see, Innis et al, PCR Strategies Academic Press, N.Y. (1995)). PCR reactions using these primers can result in the amplification of the sequence flanked by the primers. The amplified sequence can then be  
10 sequenced using methods known in the art. (see, Sambrook et al, supra, (1989)). The splice donor embodiment of the present invention are particularly useful in this regard.

Alternatively, for the reverse transcription reaction, the oligo-dT primer can have an oligo linker of known sequence. PCR can then be used to amplify the target sequence using a primer that corresponds at least to a portion of the linker and a primer that  
15 corresponds at least to a portion of the vector or the target gene.

Furthermore, nested PCR can be used with 3' RACE to enhance the sensitivity of these methods to enhance the identification of genes that are in low abundance in the target cell (for nested PCR, see Loh, Methods 2:11 (1991)). Furthermore, 5' RACE can be used to identify sequences following methods known in the art (see, EP 0731169 to  
20 Skarnes, published September 11, 1996; and Skarnes et al., Genes Dev. 6:903-918 (1992)).

### *Substrates for Measuring Beta-lactamase Activity*

25 Any membrane permanent beta-lactamase substrate capable of being measured inside the cell after cleavage can be used in the methods and compositions of the invention. Membrane permanent beta-lactamase substrates will not require permeablizing eukaryotic cells either by hypotonic shock or by electroporation. Generally, such non-specific pore forming methods are not desirable to use in eukaryotic  
30 cells because such methods injure the cells, thereby decreasing viability and introducing additional variables into the screening assay (such as loss of ionic and biological contents

of the shocked or porated cells). Such methods can be used in cells with cell walls or membranes that significantly prevent or retard the diffusion of such substrates. Preferably, the membrane permeant beta-lactamase substrates are transformed in the cell into a  $\beta$ -lactamase substrate of reduced membrane permeability (usually at least five less permeable) or that is membrane impermeant. Transformation inside the cell can occur via intracellular enzymes (e.g. esterases) or intracellular metabolites or organic molecules (e.g. sulfhydryl groups). Preferably, such substrates are fluorescent. Fluorescent substrates include those capable of changes, either individually or in combination, of total fluorescence, excitation or emission spectra or FRET.

Preferably, FRET type substrates are employed with the methods and compositions of the invention. Including fluorogenic substrates of the general formula I:



wherein D is a FRET donor and A is a FRET acceptor and S is a substrate for a protein with beta-lactamase activity. Beta-lactamase activity cleaves either D-S or S-A bonds thereby releasing either D or A, respectively from S. Such cleavage resulting from beta-lactamase activity dramatically increases the distance between D and A which usually causes a complete loss in energy transfer between D and A. Generally, molecules of D-S-A structure are constructed to maximize the energy transfer between D and A.

Preferably, the distance between D and A is generally equal to or less than the  $R_0$ .

As would readily be appreciated by those skilled in the art, the efficiency of fluorescence resonance energy transfer depends on the fluorescence quantum yield of the donor fluorophore, the donor-acceptor distance and the overlap integral of donor fluorescence emission and acceptor absorption. The energy transfer is most efficient when a donor fluorophore with high fluorescence quantum yield (preferably, one approaching 100%) is paired with an acceptor with a large extinction coefficient at wavelengths coinciding with the emission of the donor. The dependence of fluorescence energy transfer on the above parameters has been reported Forster, T. (1948) *Ann. Physik* 2: 55-75; Lakowicz, J. R., *Principles of Fluorescence Spectroscopy*, New York: Plenum Press (1983); Herman, B., Resonance energy transfer microscopy, in: *Fluorescence Microscopy of Living Cells in Culture, Part B, Methods in Cell Biology*, Vol. 30, ed. Taylor, D.L. & Wang, Y.L., San Diego: Academic Press (1989), pp. 219-243; Turro, N.

J., *Modern Molecular Photochemistry*, Menlo Park: Benjamin/Cummings Publishing Co., Inc. (1978), pp. 296-361, and tables of spectral overlap integrals are readily available to those working in the field for example, Berlman, I.B. *Energy transfer parameters of aromatic compounds*, Academic Press, New York and London (1973). The distance  
5 between donor fluorophore and acceptor dye at which fluorescence resonance energy transfer (FRET) occurs with 50% efficiency is termed  $R_0$  and can be calculated from the spectral overlap integrals. For the donor-acceptor pair fluorescein - tetramethyl rhodamine which is frequently used for distance measurement in proteins, this distance  $R_0$  is around 50-70 Å dos Remedios, C.G. et al. (1987) *J. Muscle Research and Cell*  
10 *Motility* 8:97-117. The distance at which the energy transfer in this pair exceeds 90% is about 45 Å. When attached to the cephalosporin backbone the distances between donors and acceptors are in the range of 10 Å to 20 Å, depending on the linkers used and the size of the chromophores. For a distance of 20 Å, a chromophore pair will have to have a calculated  $R_0$  of larger than 30 Å for 90% of the donors to transfer their energy to the  
15 acceptor, resulting in better than 90% quenching of the donor fluorescence. Cleavage of such a cephalosporin by beta-lactamase relieves quenching and produces an increase in donor fluorescence efficiency in excess of tenfold. Accordingly, it is apparent that identification of appropriate donor-acceptor pairs for use as taught herein in accordance with the present invention would be essentially routine to one skilled in the art.

20 Reporting gene substrates described in Tsien et al., PCT Publication No. WO96/30540 published October 3, 1996 are preferred for beta-lactamase.

### *Fluorescence Measurements*

25 When using fluorescent substrates, it will be recognized that different types of fluorescent monitoring systems can be used to practice the invention. Preferably, FACS systems are used or systems dedicated to high throughput screening e.g., 96 well or greater microtiter plates. Methods of performing assays on fluorescent materials are well known in the art and are described in, e.g., Lakowicz, J. R., *Principles of Fluorescence*  
30 *Spectroscopy*, New York: Plenum Press (1983); Herman, B., Resonance energy transfer microscopy, in: *Fluorescence Microscopy of Living Cells in Culture, Part B, Methods in*

*Cell Biology*, vol. 30, ed. Taylor, D.L. & Wang, Y. L., San Diego: Academic Press (1989), pp. 219-243; Turro, N. J., *Modern Molecular Photochemistry*, Menlo Park: Benjamin/Cummings Publishing Col, Inc. (1978), pp. 296-361.

Fluorescence in a sample can be measured using a fluorimeter. In general, excitation radiation, from an excitation source having a first wavelength, passes through excitation optics. The excitation optics cause the excitation radiation to excite the sample. In response, fluorescent proteins in the sample emit radiation that has a wavelength that is different from the excitation wavelength. Collection optics then collect the emission from the sample. The device can include a temperature controller to maintain the sample at a specific temperature while it is being scanned. According to one embodiment, a multi-axis translation stage moves a microtiter plate holding a plurality of samples in order to position different wells to be exposed. The multi-axis translation stage, temperature controller, auto-focusing feature, and electronics associated with imaging and data collection can be managed by an appropriately programmed digital computer. The computer also can transform the data collected during the assay into another format for presentation.

Preferably, FRET is used as a way of monitoring beta-lactamase activity inside a cell. The degree of FRET can be determined by any spectral or fluorescence lifetime characteristic of the excited construct, for example, by determining the intensity of the fluorescent signal from the donor, the intensity of fluorescent signal from the acceptor, the ratio of the fluorescence amplitudes near the acceptor's emission maxima to the fluorescence amplitudes near the donor's emission maximum, or the excited state lifetime of the donor. For example, cleavage of the linker increases the intensity of fluorescence from the donor, decreases the intensity of fluorescence from the acceptor, decreases the ratio of fluorescence amplitudes from the acceptor to that from the donor, and increases the excited state lifetime of the donor.

Preferably, changes in the degree of FRET are determined as a function of the change in the ratio of the amount of fluorescence from the donor and acceptor moieties, a process referred to as "ratioing." Changes in the absolute amount of substrate, excitation intensity, and turbidity or other background absorbances in the sample at the excitation wavelength affect the intensities of fluorescence from both the donor and acceptor

approximately in parallel. Therefore the ratio of the two emission intensities is a more robust and preferred measure of cleavage than either intensity alone.

The excitation state lifetime of the donor moiety is, likewise, independent of the absolute amount of substrate, excitation intensity, or turbidity or other background  
5 absorbances. Its measurement requires equipment with nanosecond time resolution, except in the special case of lanthanide complexes in which case microsecond to millisecond resolution is sufficient.

The ratio-metric fluorescent reporter system described herein has significant advantages over existing reporters for gene integration analysis, as it allows sensitive  
10 detection and isolation of both expressing and non-expressing single living cells. This assay system uses a non-toxic, non-polar fluorescent substrate that is easily loaded and then trapped intracellularly. Cleavage of the fluorescent substrate by beta-lactamase yields a fluorescent emission shift as substrate is converted to product. Because the beta-lactamase reporter readout is ratiometric it is unique among reporter gene assays in that it  
15 controls for variables such as the amount of substrate loaded into individual cells. The stable, easily detected, intracellular readout eliminates the need for establishing clonal cell lines prior to expression analysis. With the beta-lactamase reporter system or other analogous systems flow sorting can be used to isolate both expressing and non-expressing cells from pools of millions of viable cells. This positive and negative selection allows  
20 its use with gene identification methods to isolate desired clones from large clone pools containing millions of cells each containing a unique integration site.

#### *High Throughput Screening System*

25 The present invention can be used with systems and methods that utilize automated and integratable workstations for identifying modulators, pathways, chemicals having useful activity and other methods described herein. Such systems are described generally in the art (see, U.S. Patent Nos: 4,000,976 to Kramer et al. (issued January 4, 1977), 5,104,621 to Pfoest et al. (issued April 14, 1992), 5,125,748 to Bjornson et al.  
30 (issued June 30, 1992), 5,139,744 to Kowalski (issued August 18, 1992), 5,206,568 Bjornson et al. (issued April 27, 1993), 5,350,564 to Mazza et al. (September 27, 1994),



5,589,351 to Harootunian (issued December 31, 1996), and PCT Application Nos: WO 93/20612 to Baxter Deutschland GMBH (published October 14, 1993), WO 96/05488 to McNeil et al. (published February 22, 1996) and WO 93/13423 to Agong et al. (published July 8, 1993).

Typically, such a system includes: A) a storage and retrieval module comprising storage locations for storing a plurality of chemicals in solution in addressable wells, a well retriever and having programmable selection and retrieval of the addressable wells and having a storage capacity for at least 10,000 the addressable wells, B) a sample distribution module comprising a liquid handler to aspirate or dispense solutions from selected the addressable wells, the chemical distribution module having programmable selection of, and aspiration from, the selected addressable wells and programmable dispensation into selected addressable wells (including dispensation into arrays of addressable wells with different densities of addressable wells per centimeter squared), C) a sample transporter to transport the selected addressable wells to the sample distribution module and optionally having programmable control of transport of the selected addressable wells (including adaptive routing and parallel processing), D) a reaction module comprising either a reagent dispenser to dispense reagents into the selected addressable wells or a fluorescent detector to detect chemical reactions in the selected addressable wells, and. a data processing and integration module. The addressable wells should be made of biocompatible materials that are also compatible with the assay to be performed (see, U.S. Patent Application Attorney Docket No.: 08366/008001, "Systems and methods for rapidly identifying useful chemicals in liquid samples" (Stylli et al., filed May 16, 1997), which is incorporated herein by reference.

The storage and retrieval module, the sample distribution module, and the reaction module are integrated and programmably controlled by the data processing and integration module. The storage and retrieval module, the sample distribution module, the sample transporter, the reaction module and the data processing and integration module are operably linked to facilitate rapid processing of the addressable sample wells. Typically, devices of the invention can process about 10,000 to 100,000 addressable wells, which can represent about 5,000 to 100,000 chemicals, in 24-hour period. Cells clones generated using the present invention can be individually deposited into wells of a

multi-well platform having any number of wells, such as 96, 864, 3456, or more. The cells in the wells can be cultured, stored, screened, and inventoried using such a system.

The present invention is also directed to chemical entities and information (e.g., modulators or chemicals or databases biological activities of chemicals or targets)

5 generated or discovered by operation of the present invention, particularly chemicals and information generated using such systems.

*Pharmacology, Toxicity, Efficacy, Selectivity of Candidate Modulators*  
.....

10 The pharmacology, toxicity, efficacy and selectivity of candidate modulators can be determined using methods known and recognized in the art, such as those described in PCT/US97/17395 to Whitney et al., filed September 25, 1997.

*Compositions*

15 The present invention also encompasses a modulator in a pharmaceutical composition comprising a pharmaceutically acceptable carrier prepared for storage and subsequent administration, which have a pharmaceutically effective amount of the candidate modulator in a pharmaceutically acceptable carrier or diluent. Chemicals  
20 identified by the methods described herein do not include chemicals publicly available as of the filing date of the present application or in the prior art. Acceptable carriers or diluents for therapeutic use are well known in the pharmaceutical art, and are described, for example, in Remington's Pharmaceutical Sciences, Mack Publishing Co. (A.R. Gennaro edit. 1985). Preservatives, stabilizers, dyes and even flavoring agents may be  
25 provided in the pharmaceutical composition. For example, sodium benzoate, sorbic acid and esters of p-hydroxybenzoic acid may be added as preservatives. In addition, antioxidants and suspending agents may be used. The compositions of the present invention may be formulated and used using methods and compounds as is known in the art, such as those described in PCT/US97/17395 to Whitney et al., filed September 25,  
30 1997.

## EXAMPLES

### Example 1 Beta-lactamase Expression Constructs

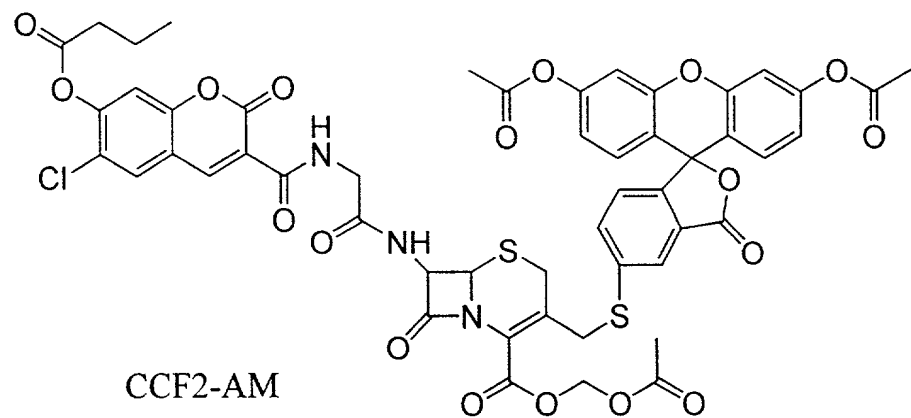
To investigate various beta-lactamase expression constructs (**BLECs**) multiple **BLECs** were constructed and transfected into mammalian cells.

The first of these, **BLEC-1** was constructed by cloning the cytoplasmic form of beta-lactamase **SEQ.ID NO. 4** (see **Table 1**) such that it is functionally linked to the En-2 splice acceptor sequence, as shown in **FIG. 3A**. This vector when inserted into a genomic intron will result in the generation of a fusion RNA between an endogenous target gene and beta-lactamase ("BL"). **BLEC-1** also contains a bovine growth hormone poly-adenylation sequence (BGH-polyA) downstream of the cytoplasmic beta-lactamase (see **Table 2**).

**BLEC-2** was constructed identically to **BLEC-1**, except that a poliovirus internal ribosomal entry site (IRES) sequence was inserted between the En-2 splice acceptor beta-lactamase ("BL"). This eliminates reading frame restrictions and possible inactivation of beta-lactamase by fusion to an endogenous protein. To allow for selection of stable transfectants for **BLEC-1** and **BLEC-2** a neomycin or G418 resistance cassette was cloned downstream of the BGH poly-adenylation sequence. This cassette sequence comprises a promoter, neomycin resistance gene and an SV40 poly-adenylation sequence, as shown in **FIG. 3A**. A version of these plasmid constructs can be inserted into retroviral vectors. One example of such constructs is shown in **FIG. 3B**.

Two alternative constructs **BLEC-3** and **BLEC-4** were constructed similar to **BLEC-1**, and **BLEC-2** respectively, except the SV40-poly A was replaced with a splice donor sequence (see, **Table 2**). This should enrich for insertion into transcribed regions, as it requires the presence of an endogenous splice acceptor and polyadenylation sequence downstream of the vector insertion site to generate G418 resistant clones. **BLEC-3** and **BLEC-4** also use the PGK promoter to drive the neomycin resistance gene instead of the human beta-actin promoter.

The structure of CCF2/AM (BL substrate) used in the experiments below is:



5

Table 1

SEQ. ID NO.	parent -BL gene and reference	Modification	mammalian expression vector	Location of expression
# 1	<i>Escherichia coli</i> RTEM Kadonaga et al.	Signal sequence replaced by: <b>ATG AGT</b>	pMAM-neo glucocorticoid-inducible	Cytoplasmic
# 2	<i>Escherichia coli</i> RTEM Kadonaga et al.	Wild type secreted enzyme 2 changes in pre-sequence: ser 2 arg , ala 23 gly	pMAM-neo glucocorticoid-inducible	Secreted extracellularly
# 3	<i>Escherichia coli</i> RTEM	-globin up stream leader: AAGCTTTTTGCAGAAGCTCA GAATAAACGCAACTTTCCG Kozak sequence: GGTACCACCATGG signal sequence replaced by: <b>ATG GGG</b>	pCDNA 3 CMV promotor and pZEO SV40 promotor	Cytoplasmic
# 4	<i>Escherichia coli</i> RTEM	Kozak sequence: GGTACCACCATGG signal sequence replaced by: <b>ATG GAC</b> (GAC replaces CAT)	pCDNA3 CMV promoter AND BLECs	Cytoplasmic
# 5	<i>Bacillus licheniformis</i> 749/C Neugebauer et al.	signal sequence removed, new N-terminal ATG	pCDNA 3 CMV promotor	Cytoplasmic

10

Table 2

Functional Elements

	Splice acceptor	Adapter	Reporter gene	Reporter gene Poly A	Selection Promoter	Resistant Marker poly A
VECTORS						
BLEC-1	En2-splice acceptor	protein fusion	SEQ. ID NO. 4	BGH polyA	β-actin promoter	Neo polyA
BLEC-2	En2-splice acceptor	IRES	SEQ. ID NO. 4	BGH polyA	β-actin promoter	Neo- PolyA
BLEC-3	En2-splice acceptor	Protein fusion	SEQ. ID NO. 4	BGH polyA	PGK promoter	Neo- Splice Donor
BLEC-4	En2-splice acceptor	IRES	SEQ. ID NO. 4	BGH polyA	PGK promoter	Neo- Splice Donor

Example 2 Libraries of BLEC Clones

To investigate the function of each of the BLEC vectors they were transfected by electroporation into RBL-1 cells and stable clones were selected for each of the four BLEC plasmids (see Table 2). Selective media contained DMEM, 10% fetal bovine serum (FBS) and 400 µg/ml Geneticin (G418). G418 resistant cell clones were pooled from multiple transfections to generate a library of BLEC stable integrated clones.

This library of BLEC-1 integrated clones was loaded with the fluorescent substrate of BL (CCF-2-AM) by adding 10 microM CCF-2-AM in HBSS containing 10 microM HEPES at pH 7. 1 and 1% glucose. After a 1 hour incubation at 22°C cells were washed with HBSS and viewed upon excitation with 400nm light using a 435nm long pass emission filter. Under these assay conditions 10% of the cells were blue fluorescent indicating they were expressing beta-lactamase. This result suggests that that BLEC-1 construct is functioning as a gene integration vector.

Stable cell lines were also generated by transfecting BLEC-1 into CHO-K1 and Jurkat cells. Populations of BLEC-1 integrated clones from CHO and Jurkat cells showed similar results to those obtained with RBL-1 clones with 10-

15 % of **BLEC** integrated cell clones expressing **BL** as determined by their blue/green ratio after loading with CCF-2-AM. This result shows that the **BLECs** function in a variety of cell types including human T-cells (Jurkat), rat basophilic leukocytes (RBL), and Chinese hamster ovarian (CHO).

5

### Example 3 Isolating BLEC Clones Expressing $\beta$ -lactamase

Fluorescent activated cell sorting of multi-clonal populations of RBL-1 gene integrated clones was used to identify clones with regulated **BL** gene expression. A **BL** non-expressing population of cells was isolated by sorting a library of **BLEC-1** integrated clones generated by transfection of RBL-1 cells as described in **Example 2**. 180,000 clones expressing little or no **BL** were isolated by sorting for clones with a low blue/green ratio (R1 population), as shown in **FIG. 4A**. This population of clones was grown for seven days and resorted by FACS to test the population's fluorescent properties. FACS analysis of the cell clones sorted from R1 shows that most of the cells with a high blue/green ratio ~0.1% have been removed by one round of sorting for green cells, as shown in **FIG. 4B**. It is also clear that the total population has shifted towards more green cells compared to the parent population, as shown in **FIG. 4A**. There are, however, cells with a high blue/green ratio showing up in the green sorted population. These may represent clones in which the **BLEC** has integrated into a differentially regulated gene such as a gene whose expression changes throughout the cell cycle.

The population of RBL-1 clones shown in **FIG. 4B** was stimulated by addition of 1uM ionomycin for 6 hours and resorted to identify clones which had the **BLEC** integrated into a gene which is inducible by increasing intracellular calcium. **Table 3** below summarizes the results from this experiment. A greater percentage of blue clones were present in all three of the blue sub-population (R4, R2, R5) in the ionomycin stimulated when compared to the unstimulated population. This sorted population represents the following classes of blue cells: R4 (highest blue/green ratio (bright blues)), R2 (multicolor blues), and R5 (lower

blue/green ratio (least blue). Additionally, in the ionomycin stimulated population there is a decrease in the percent green cells from the unstimulated population (R6). This increase in blue clones in the ionomycin stimulated population indicates that a sub-population of blue clones have the **BLEC** inserted into a gene which is induced by ionomycin. Individual blue clones were sorted from the ionomycin stimulated population and are analyzed for their expression profile.

**Table 3**  
**Sort Window (See FIG. 4)**

	R4 (blue)	R2	R5	R6 (green)
Unstimulated %	.11	2.39	1.53	66.23
1uM Ionomycin Stimulated %	.24	3.5	2.5	61.64
Ratio +Ion/-Ion	2.2	1.5	1.6	.9

In addition to allowing the isolation of cell clones with inducible BL expression from large populations of cells, clones can be isolated based their level of **BL** expression. To isolate cells with different levels of **BL** expressions blue clones can be sorted after different exposure times to substrate or by their blue/green ratio. Cell with a lower blue/green ratio or those requiring longer incubation times will represent clones expressing lower levels of **BL**. This is demonstrated by the FACS scan above as clones sorted from the R4 window have a higher blue/green ration indicating they are expressing higher levels of **BL**, cells sorted from the R5 have a lower blue/green ratio (visually turquoise) indicating lower **BL** expression. Cell sorted from the R3 window which contain all the blue cells show variation in blue color from bright blue (high blue/green ratio) to turquoise blue (low blue/green ratio).

To demonstrate that the expression constructs are relatively stable for sorted clones cells were sorted from R3 (blue population) as shown in **FIG. 4A** and cultured in the absence of selective pressure for several weeks. There was little change in the percent of blue cells in the cultured population with the percent

blue being maintained at ~90%. This result represents a 10-fold enrichment for clones constitutively expressing **BL** by one round of FACS selection.

Cells in R6 window have the lowest blue/green ration and appear green visually. R6 cell is therefore not expressing **BL** or are expressing **BL** below the  
5 detection limit of our assay.

#### Example 4 Stability of BLEC Clones

To further investigate the stability of reporter gene integrations into  
10 constitutively active genes, single blue clones were sorted from cell clone  
populations generated by transfecting RBL-1, and CHOK1 with **BLEC-1**. After  
addition of CCF-2 to the multi-clonal cell population, single blue clones were  
sorted into 96 well microtiter plates. These clones were expanded to 24 well  
dishes that took 7-10 days. The cell viability varied between the two cell types  
15 with 80% of the sorted clones forming colonies for the CHO and 36% for the  
RBL-1 cells. After expansion into a 24 well dishes 20 CHO **BLEC-1** stable  
clones were tested for **BL** expression by addition of CCF-2 -AM. 20/20 of these  
clones expressed **BL** with the percent blue cells within a clone ranging from 70%  
to 99%. This result is consisted with the earlier data presented for RBL-1 in which  
20 the blue sorted population was tested for **BL** expression after several weeks of  
non-selective culturing. There was however a significant differences between  
clones in their blue/green ratio and hence their level of **BL** expression. This  
suggested that genes with different levels of constitutive expression had been  
tagged with the **BLEC**. Although there was a significant differences in blue color  
25 between separate clones the blue fluorescence within a clone was consistently  
similar as would be expected in a clonal population. There were however green  
cells within the blue sorted clones, which may indicate that there is some loss of  
the **BLEC-1** plasmid integration site when clones are grown up from a single cell.

Single clones were expanded and used to make RNA for RACE to identify  
30 the target gene and DNA for southern analysis.



### Example 5 Isolation of Jurkat BLEC integrated clones that constitutively express beta-lactamase

Jurkat cells are a T-cell line derived from a human T-cell leukemia. This cell line maintains many of the signaling capabilities of primary T-cells and can be activated using anti-CD3 antibodies or mitogenic lectins such as phytohemagglutinin (PHA). Wild type Jurkat cells were transfected by electroporation with a beta-lactamase trapping construct (**BLEC-1**, **BLEC-1A**, or **BLEC-1B** see **FIG. 3**) (“**BLEC** constructs”) that contains a gene encoding an beta-lactamase gene that is not under control of a promoter recognized by the Jurkat cells and a neomycin resistance gene that can be expressed in Jurkat cells. **BLEC-1** is set forth in **FIG. 3**. **BLEC-1A** has a NotI site after the SV40 poly A site. This allows the cutting of the insert away from the plasmid backbone. **BLEC-1B** is the same as **BLEC-1A** except that the ATG at the beta-lactamase translation start has been changed to ATC. This eliminated the translation start site and requires the addition of an upstream ATG to produce beta-lactamase. Stable transformants were selected for their resistance to 800 micrograms/ml G418. After 400 separate experiments, a pool of greater than one million clones with **BLEC** insertions was produced. This population of cells is a library of cell clones in which the **BLEC** construct inserted throughout the genome (“Jurkat **BLEC** library”). Approximately ten percent of the cells in this library express beta-lactamase in the absence of added stimuli. Beta-lactamase activity in the cells was determined by contacting the cells with CCF2/AM and loading in the presence of Pluronic 128 (from Sigma) at a about 100 micrograms/ml. Individual clones or populations of cells that express beta-lactamase can be obtained by FACS sorting.

Genomic Southern analysis of these clones using a DNA probe encoding beta-lactamase showed the vector inserted into the host genome between one and three times per cell, with most clones having one or two vector insertion sites (for Genomic Southern analyses, see Sambrook, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press (1989)). Northern analysis of these clones using a DNA probe that encodes beta-lactamase showed that the level of expression and message size varied from clone to clone (for Northern analysis, see Sambrook, supra, (1989)). This indicated that

fusion transcripts were being made with different genes functionally tagged with beta-lactamase, which allows for the reporter gene to be expressed under the same conditions as the endogenous gene. Using appropriate primers, RACE (Gibco BRL) was used to isolate the genes linked to the expressed beta-lactamase gene in a subset of these  
5 constitutively expressing clones. These genes were cloned and sequenced using known methods (see, Sambrook, supra, (1989)). These sequences were compared with known sequences using established BLAST search techniques. Known sequences that were identified included: beta-catenin, moesin, and beta-adaptin. Additionally, several novel sequences were identified which represent putative genes.

#### Example 6 Isolation of Jurkat BLEC integrated clones that show induced expression of beta-lactamase upon activation

Jurkat BLEC integrated clones that exhibit beta-lactamase expression upon  
15 activation of the Jurkat cells by PHA (PHA induced clones) were isolated by FACS sorting a Jurkat BLEC library. These clones represent cells in which the trapping construct had integrated into a gene up regulated by PHA (T-cell) activation. Thus, these cells report the transcriptional activation of a gene upon cellular activation. Individual clones were identified and isolated by FACS using CCF2/AM to detect beta-lactamase  
20 activity. This clone isolation method, the induced sorting paradigm, used three sequential and independent stimulation and sorting protocols. A FACS read out for Jurkat cells that do not contain a BLEC construct contacted with CCF2/AM was used as a control. These control cells were all green.

The first sorting procedure isolated a pool of blue (beta-lactamase expressing, as  
25 indicated by contacting the cells with CCF2/AM) clones which had been pre-stimulated for 18 hours with 10 microgram/ml PHA from an unsorted Jurkat BLEC library. This pool represented 2.83 % of the original unsorted cell population. This selected pool contained clones that constitutively express beta-lactamase and clones in which the beta-lactamase expression was induced by PHA stimulation ("stimulatable clones"). After  
30 sorting, this pool of clones was cultured in the absence of PHA to allow the cells, in the

case of stimutable clones, to expand and return to a resting state (i.e. lacking PHA induced gene expression).

The second sorting procedure isolated a pool of green (non- $\beta$ -lactamase expressing, as indicated by contacting the cells with CCF2/AM) cell clones from the first sorted pool that had been grown, post-sorting, without PHA stimulation for 7 days. The second sorting procedure separates clones that constitutively express beta-lactamase from cells that express beta-lactamase upon stimulation. This second pool represented 11.59% of the population of cells prior to the second sort. This pool of cells was cultured in the absence of PHA to amplify the cell number prior to a third sort.

The third sorting procedure used the same procedure as the first sorting procedure and was used to isolate individual cells that express beta-lactamase in response to being contacted with 10 micrograms/ml PHA for 18 hours. Single blue clones were sorted individually into single wells of 96 well microtiter plates. This three round FACS sorting procedure enriched PHA inducible clones about 10,030 fold.

These isolated clones were expanded and tested for PHA inducibility by microscopic inspection with and without PHA stimulation in the presence of CCF2/AM. A total of fifty-five PHA inducible clones were identified using this procedure. The PHA inducibility for these clones ranged from a 1.5 to 40 fold change in the 460/530 ratio as compared to unstimulated control cells. Genomic Southern analysis using a DNA probe encoding beta-lactamase established that these clones represented 34 independent stable vector integration events. A list of clones obtained by the methods of the present invention and their characteristics is provided below in **Table 6** and **Table 7**.

In addition to PHA inducible clones, Phobol 12-myristate 13-acetate (PMA) (Calbiochem), Thapsigargin (Thaps) (Calbiochem), and PMA + Thaps inducible clones were isolated using the general procedure set forth above using the indicated inducer rather than PHA. PMA is a specific activator of PKC (protein kinase C) and Thaps is a specific activator of intracellular calcium ion release (Thaps). These clones were isolated using three rounds of FACS using the general procedures described for the PHA inducible clones in Example 5. In such instances, other stimulants were substituted for PHA. PMA was provided at 8 nM, Thaps was provided at 1 microM. When these two stimulants were combined, their concentration was not changed. As shown in **Table 5**,

clones were selected based on their activation by PMA, Thaps, or PMA with Thaps after three or eighteen hours of stimulation ("stimulation time"). These results demonstrate that the FACS sorting criteria can be varied depending upon the type of modulated clones desired. By using varied selection conditions, it is possible to isolate functionally distinct clones downstream of the desired signaling target.

### **Example 7 Isolation of Jurkat BLEC integrated clones that show repressed expression of beta-lactamase upon activation**

Jurkat **BLEC** clones that exhibit decreased beta-lactamase expression upon activation of the Jurkat cells by PHA were isolated by FACS sorting. These clones represent cells in which the BLEC trapping construct had integrated into a gene down regulated by PHA (T-cell) activation. Thus, these cells report the transcriptional repression of a gene upon cellular activation. Individual clones were identified and isolated by FACS using CCF2/AM to detect beta-lactamase activity using the following repressed sorting paradigm.

A first sort was used to isolate a population of cells that constitutively express beta-lactamase by identifying and isolating a population of blue cells from an unstimulated population of **BLEC** transfected Jurkat cells contacted with CCF2/AM. The sorted population of cells represented 2.89% of the unsorted population. These cells were cultured, divided into two pools, and stimulated with one of two different stimuli, either 10 micrograms/ml PHA for 18 hours, or 8 nM PMA and 1 microM Thapsigargin for 18 hours. These stimulated cells were contacted with CCF2 (loading in the presence of 400 PET (4% weight/volume) and Pluronic 128 (100 micrograms/ml)) and the green cells in the population were sorted using FACS. The sorted population represented 8.41 % of the cell population prior to the second sort. The third round of FACS was for single blue unstimulated cells. The population of cells obtained represented 18.2 % of the cell population prior to the third sort.

This sorting procedure represents a 2,260-fold enrichment for PHA repressible clones. These clones have the beta-lactamase gene integrated into a gene that is down regulated by PHA stimulation of the cells. Six of 80 individual clones tested were

repressed by PHA or PMA + Thapsigargin. All of these clones were confirmed to be independent integration events by genomic Southern analysis using a DNA probe encoding beta-lactamase. The results of these studies are presented in **Table 5**.

**TABLE 4**

**Identification of trapping cell lines with reporter genes  
expression which is regulated by T-cell activation**

Stimuli (Dose)	First Sort Activation Chemical and Time of Exposure	Stimulation Time	Sorting Paradigm	Clones Isolated	Clones with One or Two Vector Insertion(s)	
					1	2
PHA (10 micrograms/ml)	PHA 18 hours	18 hours	Induced	34	24	10
PMA (8 nM) + Thaps (1 microM)	PMA + Thaps 3 hours	3 hours	Induced	2	2	0
PMA (8 nM)	PMA 3 hours	3 hours	Induced	3	2	1
Thaps (1 microM)	Thaps 3 hours	3 hours	Induced	2	2	0
PHA (10 micrograms/ml) or PMA (8nm) + Thaps (1 microM)	No Stimulation	18 hours	Repressed	6	5	1

### Example 8 Specificity of T-cell modulated clones

Isolated clones from PHA-induced (Example 6) and PHA-repressed (Example 7) procedures described above were characterized to determine the specificity of their modulation and time required for induction or repression. Clones were stimulated with multiple activators or inhibitors over a one to twenty-four hour time interval. As shown in **Table 6**, five clones produced by the induced and repressed sorting paradigms using a plurality of activators were tested for their responsiveness to a variety of T-cell activators, suppressors, and combinations thereof.

**TABLE 6**  
**Sorting protocols and specificity of activated BLEC Jurkat clones**

Clone	Sorting Procedures				Relative Beta-Lactamase Activity of the Clone by the Indicated Stimulus After 24 hours (% of maximum activated stimuli)						
	Paradigm	First Sort Stimulus and (cell color sorted for)	Second Sort Stimulus and (cell color sorted for)	Third Sort Stimulus And (cell color sorted for)	None	PMA (8nM)	Thaps (1micro M)	PMA (8nM) + Thaps (1micro M)	PMA (8nM) + Thaps (1micro M) + CsA (100 nM)	PHA (10 microgr am/ml)	PHA (10micr ogram/ml) + CsA (100 nM)
J83-PI9	Induced	PHA <sup>a</sup> (blue)	N/S (green)	PHA (blue)	0	<1	100	50	<5	60	<5
J32-6D4	Induced	PHA (blue)	N/S (green)	PHA (blue)	0	60	1-2	100	70	80	75
C2	N/S	N/S	N/S	N/S	0	<1	0	100	<1	30	1
J389-PT14	Induced	PMA <sup>b</sup> + Thaps <sup>c</sup> (blue)	N/S (green)	PMA + Thaps (blue)	0	90	5	85	100	85	90
J83 97-PPTR2	Repressed	N/S (blue)	PMA + Thaps (green)	N/S (blue)	0	100	85	-50	85	67	75
J83-PT18	Induced	PHA (blue)	N/S (green)	PHA (blue)	0	80	100	25	70	60	60

"N/S" means "no stimulation"

<sup>a</sup> concentration of PHA used was 10 microgram/ml.

<sup>b</sup> concentration of PMA used was 8 nM.

<sup>c</sup> concentration of Thaps used was 1 microM.

In this study, PMA, which is a PKC activator, Thapsigargin which increases intracellular calcium, PHA which activates the T-cell receptor pathway, and cyclosporin A which is a clinically approved immunosuppressant that inhibits the Ca<sup>2+</sup> dependent phosphates calcineurin were investigated for their ability to modulate beta-lactamase expression in PHA induced and repressed **BLEC** clones.

The selected clones show varied dependence for their activation and inhibition by these activators and inhibitors which give and indication of the signaling events required for their transcriptional activation. Five of the listed clones were generated using the approaches described above in Example 6. The clone C2 was generated using a more classical approach. This clone was generated by transfecting a plasmid construct in which a 3X NFAT response element has been operably linked to beta-lactamase expression.

This 3XNFAT element represents a DNA sequence that is present in the promoter region of IL-2 and other T-cell activated genes. In addition the C2 cell line has been stably transfected with the M1 muscurinic receptor. This allows the activation of beta-lactamase expression in this clone using an M1-muscurinic agonist such as carbachol. This cell line therefore represents a good control for the cellular activators and inhibitors tested as the signaling events required for its activation are established.

The results of these studies indicate that the cell lines generated vary in their specificity towards activation or repression by activators. Thus, depending on the type of system that these cells are to be used to investigate, a panel of clones with varying specificity towards a specific pathway are made available by the present methods.

Table 7 and Table 8 provide data similar to that provided in Table 5 for all of the clones obtained by the methods of Examples 5 to 7.

**TABLE 7**  
**Characterization of induced BLEC Jurkat clones**

CLONE Number	TIME (hours) for first detectable change in color	Change in 460/530 ratio in the indicated clone by the following activator				
		PHA (10 microgram/ml)	Thaps (1 microM)	PMA (8 nM)	PMA (8 nM) + Thaps (1 microM)	Anti-CD3 (2 microgram/ml) (Pharmingen)
J325B5	6	7	Nt	2-3	Nt	4-5
J325B11	6	9	1-2	2-3	Nt	5-6
J325E3	6	7	Nt	2-3	Nt	4-5
J325G4	6	3-4	Nt	3-4	Nt	4-5
J325E6	6	11	Nt	3-4	Nt	6
J326C9	6	4-5	1-2	2-3	Nt	3-4
J325E1	<2	8	Nt	8	Nt	5-6
J326D4	<2	10	0	10	Nt	5-6
J326D7	<2	10	Nt	10	Nt	5-6
J326F7	<2	10	Nt	10	Nt	5-6
J326H4	<2	10	Nt	10	Nt	5-6
J83PI1	Nt	3-4	3-4	3-4	4-5	2-3
J83PI2	5-6	8	1-2	7-8	7-8	3-4
J83PI8	5-6	4-5	1-2	4-5	4-5	2-3
J83PI3	5-6	5-6	6-7	3-4	5-6	2-3
J83PI4	4-6	3-4	3-4	0	2-3	2
J83PI6	6-18	6-7	7-8	0	4-5	4
J83PI9	6	6	5-6	0	4-5	3-4
J83PI5	Nt	Nt	Nt	Nt	Nt	Nt
J83PI7	6-18	2	2	2	2	1.5-2
J83PI15	Nt	3-4	2	3-4	3-4	3-4
J83PI16	Nt	3-4	1-2	3-4	3-4	2-3
J83PI18	Nt	5-6	7-8	5	Nt	Nt
J83PI12	Nt	Nt	Nt	Nt	Nt	Nt
J83PI14	Nt	2	2	2	Nt	Nt
J83PI17	Nt	Nt	Nt	Nt	Nt	Nt
J83PI19	Nt	5-6	1-2	3	1-2	1-2

CLONE Number	TIME (hours) for first detectable change in color	Change in 460/530 ratio in the indicated clone by the following activator				
		PHA (10 microgram/ml)	Thaps (1 microM)	PMA (8 nM)	PMA (8 nM) + Thaps (1 microM)	Anti-CD3 (2 microgram/ml) (Pharmingen)
J83PI11	Nt	Nt	Nt	Nt	Nt	Nt
J83PI13	Nt	2-3	2-3	0	Nt	Nt
J97PI1	Nt	3-4	3-4	3-4	3-4	3-4
J97PI2	Nt	2-3	Nt	Nt	2-3	Nt
J97PI3	Nt	1-2	1-2	1-2	1-2	Nt
J97PI4	Nt	1-2	1-2	1-2	1-2	Nt
J97PI5	Nt	1-5	1-9	1-5	2-3	Nt
J97PI6	Nt	3-4	4-6	1-2	4-6	Nt
J97PI13	Nt	2-3	5-6	1-2	4-5	Nt
J97PI18	Nt	1-2	3-4	1-2	4-5	Nt
J97PI7	Nt	3-4	4-5	1-2	5-6	Nt
J97PI17	Nt	4-5	7-8	1-2	8-10	Nt
J97PI8	Nt	2-5-3	3-4	1-2	3-4	Nt
J97PI9	Nt	2-3	4-5	1-2	5-6	Nt
J97PI10	Nt	3-4	3-4	1-2	4-5	Nt
J97PI23	Nt	4-5	4-5	1-2	4-5	1-2
J97PI11	Nt	3-4	5-6	2	4-5	Nt
J97PI15	Nt	1-2	3-4	1-2	3-4	Nt
J97PI12	Nt	3-4	5-6	2-3	5-6	Nt
J97PI22	Nt	5-6	5-7	2-3	3-4	3-4
J97PI14	Nt	4-5	3-4	2	4-5	Nt
J97PI116	Nt	2-3	3-4	2-3	4	Nt
J97PI19	Nt	2-3	2-3	1-2	2-4	Nt
J97PI20	Nt	1-2	2-3	1-2	1-2	Nt
J97PI21	Nt	2-3	2-3	1-2	2-3	2-3
J97PI24	Nt	3-4	3-4	2-3	7-10	3-4
J389PT1	2hours	5-6	3-4	8-9	8-9	3-4
J389PT4	1hour	15	10	12	16	15
J389PM2	1hour	4-5	3-4	3-4	4-5	4-5
J389PM3	1hour	3-4	2-3	2-3	3-4	3-4
J389PM5	1hour	4-5	3-4	3-4	4-5	4-5
J389PM7	3hours	1-2	2-3	1-2	1-2	1-2
J389PM8	2-3hours	2-3	3-4	2-3	2-3	3-4
J389TI1	3-5hours	1-2	2-3	1-2	2-3	2-3
J389TI4	2hour	0	3-4	1-2	2-3	0

"Nt" means "not tested"



**TABLE 8**  
**Characterization of repressed BLEC Jurkat clones**

CLONE #	Relative repression of beta-lactamase in the indicated clone by the following activator			
	PHA (10 microgram/ml)	PHA (10 microgram/ml) + CsA (100 nM)	PMA (8 nM) + Thaps (1 microM)	PMA (8 nM) + Thaps (1 microM) + CsA (100 nM)
J83/97pptr1	90	90	75	75
J83/97 pptr2	10	-60	10	-80
J83/97pptr3	10	-50	10	-100
J83/97pptr4	60	60	40	70
J83/97pptr5	50	60	50	50
J83/97pptr6	70	70	70	70

To confirm that changes in reporter gene activity reflected changes in mRNA expression in these clones, Northern analysis was performed on induced, constitutive, and repressed clones using a radio labeled DNA probe directed towards the beta-lactamase gene. All clones that had beta-lactamase enzyme inducibility tested showed beta-lactamase mRNA inducibility. All clones that showed constitutive expression of beta-lactamase showed constitutive expression of beta-lactamase mRNA. All clones that showed repressed beta-lactamase expression showed repressed beta-lactamase mRNA. The message size of the control beta-lactamase mRNA was about 800 base pairs. The sizes of some from other beta-lactamase clones of the RNA were shifted higher in the gel, indicating a fusion RNA had been made between the endogenous transcript and beta-lactamase. Two known genes, CDK-6 (isolated from clone J83-PTI1) and Erg-3 (isolated from clone J89-PTI4), and two unknown genes were identified, which were isolated from clones J83PI15 and J83PI2, respectively. For clone J389-PTI4, a Northern blot was performed with the Erg-3 probe made using appropriate PCR primers determined from a published sequence which hybridizes with both the fusion RNA and the wild type RNA (for the sequence of Erg-3 see Stamminger et al., *Int. Immunol.* 5:63-70 (1993); for PCR methodologies, see U.S. Patent Nos: 4,800,159, 4,683,195, and 4,683,202). The inducibility in wild type Jurkat cells mimicked the beta-lactamase activity in this clone.

### Example 9 Screening of a library of known pharmacologically active modulators using a T-cell activated BLEC clone

T-cell clone J32-6D4 was used to identify potential inhibitors of the T-cell receptor pathway. This clone was selected for further study because it is difficult to identify chemicals that inhibit specific T-cell receptor pathway. Thus, this clone was used to identify chemicals that inhibit this T-cell receptor pathway that is also stimulated by the PKC activator PMA.

A first screen was performed using a generic set of 480 chemicals with known properties. The chemicals in this set were known to have pharmacological activity. Approximately one percent (7/480) of these chemicals showed greater than 50% inhibition of the PHA activation of beta-lactamase expression in clone J32-6D4 when tested in duplicate at 10 microM of chemical. Cells were activated with 1 microgram/ml of PHA for 18 hours in the presence of test chemicals to test for inhibitory activity. The seven chemicals that specifically inhibited clone J32-6D4 are shown in **Table 9**. Two of these chemicals specifically inhibited clone J32-6D4 and not the control C2 cell line. This assay for the specificity of inhibition included screening these 480 chemicals for inhibitory activity using clone C2, in which the M1 muscarinic receptor was linked to a NFAT beta-lactamase reporter gene readout (see Example 7). In these experiments, the inhibition measured was the inhibition of carbachol induced expression of beta-lactamase. These results, the specific inhibition of J32-6D4 cells but not C2 cells, show that the chemicals are not toxic, do not inhibit general transcription, and do not inhibit the reporter gene product.

**TABLE 9**

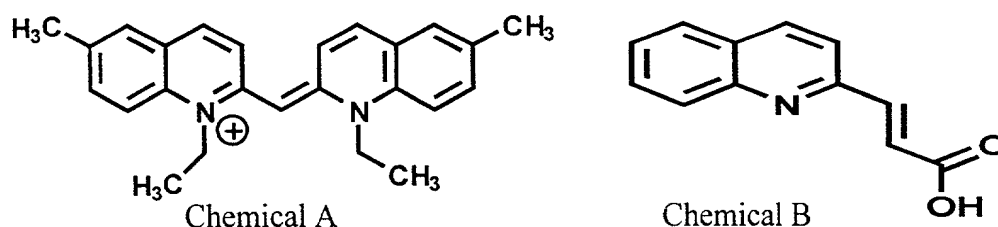
**Active chemicals identified as exhibiting inhibitory activity of PHA activation of clone J32-6D4**

Chemical (10 microM)	% Inhibition of PHA activation of Clone J32-6D4	Inhibition of Clone C2	Therapeutic Category of the Chemical
Digoxin	86	+	Cardiotonic
Digitoxin	77	+	Cardiotonic
Gentian Violet	73	+	Topical anti-infective
Oxyphenbuta zone	75	—	Anti-inflammatory
Mechloretha mine	51	—	Anti-neoplastic
Dipyrrithione	70	+	Anti-bacterial
Ouabain	50	+	Cardiotonic
Thioguanine	50	+	Anti-neoplastic

**Example 10 Screening a library of structurally characterized chemicals having unknown pharmacological properties for modulating activity of the T-cell receptor pathway using a T-cell activated BLEC clone**

Having demonstrated in Example 9 that clone J32-6D4 performs robustly in a chemical screen, this clone was used to screen an additional 7,500 chemicals from a proprietary chemical library at a concentration of 10 microM per chemical. This collection of chemicals, unlike the collection of chemicals used in Example 9, contains chemicals without known pharmacological activity. Seventy-seven chemicals showed at least 50% inhibition of PHA activation of beta-lactamase expression following the general procedures set forth in Example 7. These 77 chemicals were re-tested for this activity using the same procedure and 31 chemicals were confirmed to have activity. The IC<sub>50</sub> values of the inhibition of PHA activation of beta-lactamase expression were determined for these 31 chemicals using concentrations of chemical between about 20 microM to 2 nM. IC<sub>50</sub> values reflect the concentration of a chemical needed to inhibit the PHA activation of the clone by 50% and were determined using known methods. These 31 chemicals were also tested for their cross inhibition of cabachol induced activation of beta-lactamase expression of clone C2 as described in Example 8.

Two chemicals, designated chemical A and chemical B, exhibited an IC<sub>50</sub> values of about 200 nM and specifically inhibited the PHA activation of beta-lactamase expression of clone J32-6D4 but not the carbachol activation of clone C2 at the concentration tested. All of the other 31 chemicals either inhibited both clone J32-6D4 and clone C2, or had IC<sub>50</sub> values above 1 microM.



Chemicals A and B were further tested for their anti-proliferative effect on Jurkat cells and mouse L-cells (mouse fibroblast cell line). Chemical B showed no anti-proliferative effect on both the Jurkats and L-cells at concentrations up to 10 microM. Chemical A exhibited an anti-proliferative effect on the Jurkats and L-cells at 100 nM. Proliferation assays were performed by seeding about 20,000 cells unactivated by PHA into a 24 well plate. These cells were contacted with chemicals and were then incubated at 37°C for five days. The cells were contacted with 10 micrograms/ml of MTT (Sigma Chemical Co., MO) for three hours. The cells were then collected, resuspended in isopropanol, and the absorbance was read in a plate reader at a wavelength of 570 nM with a background subtraction at a reading at a wavelength of 690 nM (see, Carmichael et al., Cancer Res. 47:936 (1987)).

#### Example 11 Effects of identified chemicals on primary human T-cell proliferation.

An assay was developed to test the chemicals identified in Example 9 for their ability to inhibit the activation and proliferation of normal peripheral white blood cells to confirm their presumptive activity (see generally, Harlow and Lane, Antibodies, A Laboratory Manual, Cold Spring Harbor Press, (1988)). Peripheral blood from normal humans was drawn into heparanized Vacutainer® tubes and incubated with various concentrations of (superantigen) staphylococcal enterotoxin B (SEB, at 0.001 to 10 ng/ml) for 1 hour at 37°C. Brefeldin A, which was added and the cells were incubated an

additional 5 hours. EDTA was added to detach the cells, and a 100 microliter aliquot was removed, the red blood cells lysed with ammonium chloride, the remaining cells counted and their viability determined using viability staining using known methods. The red blood cells remaining in the original sample were lysed with ammonium chloride and the remaining cells (leukocytes) were permeabilized with FACS permabilizing solution using established methods. These leukocytes were harvested by centrifugation, washed and stained with the combination of antibodies CD69, IFN- $\gamma$  and CD3, which were detectably labeled. Control cells consisted of cells incubated in the absence of SEB and staining control cells consisted of cells stained with CD69/MsIgG1 and CD3 antibodies, which were detectably labeled. Similar cultures will be incubated for 71 hours, pulsed with tritiated thymidine for 1 hour and harvested and the incorporated radioactivity counted by scintillation to determine a stimulation index using established methods.

Using preferred concentrations of SEB, various concentrations of cyclosporin A (CsA) were added to determine optimal conditions of CsA for blocking of SEB stimulation of peripheral blood T-cells for use as a control for non-proliferative T-cells. Controls consisted of cells incubated with culture media in place of CsA. Control cultures incubated for 1 hour were blocked with Brefeldin A for an additional 5 hours, harvested, and stained for intracellular IFN-gamma or cultured for an additional 71 hours, pulsed with tritiated thymidine for one hour, harvested, and counted by liquid scintillation.

Using preferred concentrations of SEB and CsA, blood from normal donors was stimulated in the presence and absence of CsA. This established expected normal ranges for the degree of activation (% IFN-gamma + activated CD3+ cells for 6 hours), proliferation ( $^3\text{H}$ -TdR uptake at 72 hours) and CsA blocking at both time points.

Using preferred conditions, human blood was incubated with Chemical A or Chemical B at 2, 20, and 200 nM. CsA was used as a positive control for T-cell suppression. One hour cultures were blocked with Brefeldin A for an additional 5 hours, harvested and counted by liquid scintillation. Cell counts and percent viability were reported for each culture condition.

The results of these studies should demonstrate that at least one of the chemicals identified by the methods of the present invention have the predicted pharmacological activity in human cells.

5    **Example 12: Identification of genes expressed during developmental programs.**

Another use of this method is for the identification of genes expressed during various cellular processes, such as developmental biology and apoptosis. Genes involved in specific developmental programs, such as the differentiation of pre-adiposites to  
10    mature adiposites, can be identified using this method.

In order to practice this method, a clone library from a pre-adiposite cell line such as 3T3-L1 is made using the methods generally described in Examples 10 to 12 above. Of course, pre-adiposite cells are used rather than Jurkat cells. This cell line can be reversible differentiated to mature adipocytes by exposing them to dexamethasone and  
15    indomethasone (see, Hunt et al. Proc. Natl. Acad. Sci. U.S.A. 83:3786-3789 (1986)). These mature adiposites can be reversibly differentiated to pre-adiposites with Tumor Necrosis Factor alpha TNFa (see, Torti et al. J. Cell. Biol. 108:1105-1113 (1989)). Thus, a cell library capable of signaling the expression of genes involved in cellular differentiation can be made.

20    The 3T3-L1 gene trap library is FACS sorted to remove blue constitutively expressing beta-lactamase cells. The remaining green cells are then differentiated into mature adiposites using the dexamethasone and indomethasone. Blue (beta-lactamase expressing) cells are isolated using FACS. These clones represent cells in which the trapping construct integrates into a gene that is expressed in differentiated adiposites, but  
25    not in undifferentiated adiposites. This process can be repeated multiple times to insure enrichment for cells that express adiposite specific genes.

Alternatively, cell clones can be isolated which are differentiated for a specific time interval. For instance, blue and green cells differentiated for 2 days with dexamethasone and indomethasone are sorted. These populations of cells represent cells  
30    in which the trapping construct integrates into a gene that is expressed early in the differentiation process. This allows the identification of genes that are expressed during

the developmental program but are not expressed in pre-adiposites or mature adiposites. This method can be used to isolated genes expressed during a variety of developmental programs, including but not limited to neuronal, cardiac, muscle, and cancer cells.

These cells lines can be used to identify genes involved in the differentiation process, and can also be used to screen chemicals that modulate the differentiation process using the methods described in Examples 8 to 10 above. Drugs that can be identified include those that enhance the growth of cells, such as neuronal cells, or depress the growth or reverse differentiation of cells, such as cancer cells.

### Example 13: Assays for modulators of G-protein coupled receptors

The general procedures of Examples 8 to 10 can be used in an analogous manner to identify cell lines suitable for screens for G-protein coupled receptors (GPCRs). GPCRs are known to signal via one of several intracellular pathways. These pathways can be activated pharmacologically in cell libraries to yield potential screening cell lines. For example, Gq coupled GPCRs are known to raise intracellular free calcium via activation of phospholipase Cb (PLCb). By isolating cell lines responsive to an increase in calcium from the genomic library (e.g. induced by ionomycin or thapsigargin), screen cell lines are generated.

For example, a calcium-sensitive clone was transfected with a Gq-type GPCR by electroporation. Cells from clone J389PTI4 were transfected by electroporation with a plasmid (pcDNA3 (Invitrogen) or pcDNA3-M1 (pcDNA3 that can operably express M1 receptor) to make cell lines J389PTI4/pcDNA3 and J389PTI4/pcDNA3-M1). Cell line J389PTI4/pcDNA3-M1 expressed the M1 receptor, whereas the cell line J389PTI4/pcDNA3 did not. Thus, the J389PTI4/pcDNA3 cell is a control cell. Two days after transfection, cells were stimulated with 20 microM carbachol in 96-well microtiter plate for 6 hours in 37 °C. These cells were contacted with CCF-2 dye for another 90 minutes. The 460/530 ratio changes were measured in a Cytoflour (Series 4000 Model) (Perceptive Biosystems) fluorescence plate reader and correspond to reporter gene expression. These results are summarized in **Table 10**. The ability of the transiently-transfected clone to detect a ligand for the GPCR demonstrates the potential of generating

screening cell lines using clones made following the procedures of the present invention. The stimulation by carbachol detected in the transient tranfection assay represents a response in about 20% of the cells. To develop a stable screening cell line for the M1 receptor, this population can be sorted for individual clones responsive to carbachol and those clones can be expanded and screened to identify the most responsive clones.

Similar methods can be used to generate cell lines for Gs or Gi-coupled receptors. In these cases, clones responsive to increases or decreases in cAMP can be isolated. A variety of cell lines can be used for these procedures, such as CHO, HEK293, Neuroblastoma, P19, F11, and NT-2 cells.

**TABLE 10**

**Cell lines that report modulation of the M1 receptor pathway**

Cell Line	Relative expression of beta-lactamase in cells Exposed to the indicated stimuli		
	Unstimulated	30 $\mu$ M Carbachol	10 nM PHA
J389PTI4/pcDNA3	1	1	12
J389PTI4/pcDNA3-M1	1	4	13

## **Publications**

### ***Articles***

G. Friedrich, P. Soriano, Methods in Enzymology, Vol. 225: 681 (1993)

G. Friedrich, P. Soriano, Genes & Development, Vol. 5: 1513 (1991)

A. Gossler, et al., Reports, 28 April: 463 (1989)

D. Hill, W. Wurst, Methods in Enzymology, Vol. 225: 664 (1993)

P. Mountford, A. Smith, TIG, Vol. 11 No. 5: 179 (1995)

P. Mountford, et al., Proc. Natl. Acad. Sci, USA, Vol. 91: 4303 (1994)

I. Niwa, et al., J. Biochem, Vol. 13: 343 (1993)

U. Reddy, et al., Proc. Natl. Acad. Sci. USA, Vol. 89: 6721 (1992)

P. Shapiro, P. Senapathy, Nucleic Acids Research, Vol. 17, No. 17: 7155 (1987)

A.A. Skarnes, et al., Genes & Development, Vol. 6: 903 (1992)

W. Wurst, et al., Genetics, Vol. 139: 889 (1995)



All publications, including patent documents and scientific articles, referred to in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication were individually incorporated by reference.

All headings are for the convenience of the reader and should not be used to limit  
5 the meaning of the text that follows the heading, unless so specified.

# SEQUENCE ID. LISTING

SEQ.ID NO. 1: range 1 to 795

```

5
      10      20      30      40      50
      *      *      *      *      *
ATG AGT CAC CCA GAA ACG CTG GTG AAA GTA AAA GAT GCT GAA GAT CAG TTG
Met Ser His Pro Glu Thr Leu Val Lys Val Lys Asp Ala Glu Asp Gln Leu
10      60      70      80      90      100
      *      *      *      *      *
GGT GCA CGA GTG GGT TAC ATC GAA CTG GAT CTC AAC AGC GGT AAG ATC CTT
Gly Ala Arg Val Gly Tyr Ile Glu Leu Asp Leu Asn Ser Gly Lys Ile Leu
      110      120      130      140      150
15      *      *      *      *      *
GAG AGT TTT CGC CCC GAA GAA CGT TTT CCA ATG ATG AGC ACT TTT AAA GTT
Glu Ser Phe Arg Pro Glu Glu Arg Phe Pro Met Met Ser Thr Phe Lys Val
      160      170      180      190      200
      *      *      *      *      *
20      CTG CTA TGT GGC GCG GTA TTA TCC CGT GTT GAC GCC GGG CAA GAG CAA CTC
Leu Leu Cys Gly Ala Val Leu Ser Arg Val Asp Ala Gly Gln Glu Gln Leu
      210      220      230      240      250
      *      *      *      *      *
GGT CGC CGC ATA CAC TAT TCT CAG AAT GAC TTG GTT GAG TAC TCA CCA GTC
Gly Arg Arg Ile His Tyr Ser Gln Asn Asp Leu Val Glu Tyr Ser Pro Val
25      260      270      280      290      300
      *      *      *      *      *
ACA GAA AAG CAT CTT ACG GAT GGC ATG ACA GTA AGA GAA TTA TGC AGT GCT
Thr Glu Lys His Leu Thr Asp Gly Met Thr Val Arg Glu Leu Cys Ser Ala
30      310      320      330      340      350
      *      *      *      *      *
GCC ATA ACC ATG AGT GAT AAC ACT GCG GCC AAC TTA CTT CTG ACA ACG ATC
Ala Ile Thr Met Ser Asp Asn Thr Ala Ala Asn Leu Leu Leu Thr Thr Ile
360      370      380      390      400
35      *      *      *      *      *
GGA GGA CCG AAG GAG CTA ACC GCT TTT TTG CAC AAC ATG GGG GAT CAT GTA
Gly Gly Pro Lys Glu Leu Thr Ala Phe Leu His Asn Met Gly Asp His Val
410      420      430      440      450
      *      *      *      *      *
40      ACT CGC CTT GAT CGT TGG GAA CCG GAG CTG AAT GAA GCC ATA CCA AAC GAC
Thr Arg Leu Asp Arg Trp Glu Pro Glu Leu Asn Glu Ala Ile Pro Asn Asp
460      470      480      490      500      510
      *      *      *      *      *
GAG CGT GAC ACC ACG ATG CCT GCA GCA ATG GCA ACA ACG TTG CGC AAA CTA
Glu Arg Asp Thr Thr Met Pro Ala Ala Met Ala Thr Thr Leu Arg Lys Leu
45      520      530      540      550      560
      *      *      *      *      *
TTA ACT GGC GAA CTA CTT ACT CTA GCT TCC CGG CAA CAA TTA ATA GAC TGG
Leu Thr Gly Glu Leu Leu Thr Leu Ala Ser Arg Gln Gln Leu Ile Asp Trp
50      570      580      590      600      610
      *      *      *      *      *
ATG GAG GCG GAT AAA GTT GCA GGA CCA CTT CTG CGC TCG GCC CTT CCG GCT
Met Glu Ala Asp Lys Val Ala Gly Pro Leu Leu Arg Ser Ala Leu Pro Ala

```

```

      620      630      640      650      660
      *        *        *        *        *
GGC TGG TTT ATT GCT GAT AAA TCT GGA GCC GGT GAG CGT GGG TCT CGC GGT
Gly Trp Phe Ile Ala Asp Lys Ser Gly Ala Gly Glu Arg Gly Ser Arg Gly
5  670      680      690      700      710
      *        *        *        *        *
ATC ATT GCA GCA CTG GGG CCA GAT GGT AAG CCC TCC CGT ATC GTA GTT ATC
Ile Ile Ala Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg Ile Val Val Ile
10 720      730      740      750      760
      *        *        *        *        *
TAC ACG ACG GGG AGT CAG GCA ACT ATG GAT GAA CGA AAT AGA CAG ATC GCT
Tyr Thr Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn Arg Gln Ile Ala
      770      780      790
      *        *        *
15 GAG ATA GGT GCC TCA CTG ATT AAG CAT TGG
Glu Ile Gly Ala Ser Leu Ile Lys His Trp

```

20 SEQ.ID NO. 2: range 1 to 858

```

      10      20      30      40      50
      *        *        *        *        *
ATG AGA ATT CAA CAT TTC CGT GTC GCC CTT ATT CCC TTT TTT GCG GCA TTT
25 Met Arg Ile Gln His Phe Arg Val Ala Leu Ile Pro Phe Phe Ala Ala Phe
      60      70      80      90      100
      *        *        *        *        *
TGC CTT CCT GTT TTT GGT CAC CCA GAA ACG CTG GTG AAA GTA AAA GAT GCT
Cys Leu Pro Val Phe Gly His Pro Glu Thr Leu Val Lys Val Lys Asp Ala
30 110      120      130      140      150
      *        *        *        *        *
GAA GAT CAG TTG GGT GCA CGA GTG GGT TAC ATC GAA CTG GAT CTC AAC AGC
Glu Asp Gln Leu Gly Ala Arg Val Gly Tyr Ile Glu Leu Asp Leu Asn Ser
      160      170      180      190      200
35  *        *        *        *        *
GGT AAG ATC CTT GAG AGT TTT CGC CCC GAA GAA CGT TTT CCA ATG ATG AGC
Gly Lys Ile Leu Glu Ser Phe Arg Pro Glu Glu Arg Phe Pro Met Met Ser
      210      220      230      240      250
      *        *        *        *        *
40 ACT TTT AAA GTT CTG CTA TGT GGC GCG GTA TTA TCC CGT GTT GAC GCC GGG
Thr Phe Lys Val Leu Leu Cys Gly Ala Val Leu Ser Arg Val Asp Ala Gly
      260      270      280      290      300
      *        *        *        *        *
CAA GAG CAA CTC GGT CGC CGC ATA CAC TAT TCT CAG AAT GAC TTG GTT GAG
45 Gln Glu Gln Leu Gly Arg Arg Ile His Tyr Ser Gln Asn Asp Leu Val Glu
      310      320      330      340      350
      *        *        *        *        *
TAC TCA CCA GTC ACA GAA AAG CAT CTT ACG GAT GGC ATG ACA GTA AGA GAA
Tyr Ser Pro Val Thr Glu Lys His Leu Thr Asp Gly Met Thr Val Arg Glu
50 360      370      380      390      400
      *        *        *        *        *
TTA TGC AGT GCT GCC ATA ACC ATG AGT GAT AAC ACT GCG GCC AAC TTA CTT
Leu Cys Ser Ala Ala Ile Thr Met Ser Asp Asn Thr Ala Ala Asn Leu Leu

```

```

410          420          430          440          450
*           *           *           *           *
CTG ACA ACG ATC GGA GGA CCG AAG GAG CTA ACC GCT TTT TTG CAC AAC ATG
Leu Thr Thr Ile Gly Gly Pro Lys Glu Leu Thr Ala Phe Leu His Asn Met
5 460          470          480          490          500          510
*           *           *           *           *           *
GGG GAT CAT GTA ACT CGC CTT GAT CGT TGG GAA CCG GAG CTG AAT GAA GCC
Gly Asp His Val Thr Arg Leu Asp Arg Trp Glu Pro Glu Leu Asn Glu Ala
10          520          530          540          550          560
*           *           *           *           *           *
ATA CCA AAC GAC GAG CGT GAC ACC ACG ATG CCT GCA GCA ATG GCA ACA ACG
Ile Pro Asn Asp Glu Arg Asp Thr Thr Met Pro Ala Ala Met Ala Thr Thr
15          570          580          590          600          610
*           *           *           *           *           *
TTG CGC AAA CTA TTA ACT GGC GAA CTA CTT ACT CTA GCT TCC CGG CAA CAA
Leu Arg Lys Leu Leu Thr Gly Glu Leu Leu Thr Leu Ala Ser Arg Gln Gln
15          620          630          640          650          660
*           *           *           *           *           *
TTA ATA GAC TGG ATG GAG GCG GAT AAA GTT GCA GGA CCA CTT CTG CGC TCG
Leu Ile Asp Trp Met Glu Ala Asp Lys Val Ala Gly Pro Leu Leu Arg Ser
20          670          680          690          700          710
*           *           *           *           *           *
GCC CTT CCG GCT GGC TGG TTT ATT GCT GAT AAA TCT GGA GCC GGT GAG CGT
Ala Leu Pro Ala Gly Trp Phe Ile Ala Asp Lys Ser Gly Ala Gly Glu Arg
25          720          730          740          750          760
*           *           *           *           *           *
GGG TCT CGC GGT ATC ATT GCA GCA CTG GGG CCA GAT GGT AAG CCC TCC CGT
Gly Ser Arg Gly Ile Ile Ala Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg
25          770          780          790          800          810
*           *           *           *           *           *
ATC GTA GTT ATC TAC ACG ACG GGG AGT CAG GCA ACT ATG GAT GAA CGA AAT
Ile Val Val Ile Tyr Thr Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn
30          820          830          840          850
*           *           *           *           *           *
AGA CAG ATC GCT GAG ATA GGT GCC TCA CTG ATT AAG CAT TGG
Arg Gln Ile Ala Glu Ile Gly Ala Ser Leu Ile Lys His Trp
35

```

40 SEQ.ID NO. 3: range 1 to 795

```

AAGCTTTTTGCAGAAGCTCAGAATAAACGCAACTTTCCGGGTACCACC
45          10          20          30          40          50
*           *           *           *           *           *
*
ATG GGG CAC CCA GAA ACG CTG GTG AAA GTA AAA GAT GCT GAA GAT CAG TTG GGT
GCA
50 Met Gly His Pro Glu Thr Leu Val Lys Val Lys Asp Ala Glu Asp Gln Leu Gly
Ala
60          70          80          90          100
*           *           *           *           *           *

```

CGA GTG GGT TAC ATC GAA CTG GAT CTC AAC AGC GGT AAG ATC CTT GAG AGT  
Arg Val Gly Tyr Ile Glu Leu Asp Leu Asn Ser Gly Lys Ile Leu Glu Ser  
110 120 130 140 150  
\* \* \* \* \*  
5 TTT CGC CCC GAA GAA CGT TTT CCA ATG ATG AGC ACT TTT AAA GTT CTG CTA  
Phe Arg Pro Glu Glu Arg Phe Pro Met Met Ser Thr Phe Lys Val Leu Leu  
160 170 180 190 200 210  
\* \* \* \* \*  
TGT GGC GCG GTA TTA TCC CGT GAT GAC GCC GGG CAA GAG CAA CTC GGT CGC  
10 Cys Gly Ala Val Leu Ser Arg Ile Asp Ala Gly Gln Glu Gln Leu Gly Arg  
220 230 240 250 260  
\* \* \* \* \*  
CGC ATA CAC TAT TCT CAG AAT GAC TTG GTT GAG TAC TCA CCA GTC ACA GAA  
Arg Ile His Tyr Ser Gln Asn Asp Leu Val Glu Tyr Ser Pro Val Thr Glu  
15 270 280 290 300 310  
\* \* \* \* \*  
AAG CAT CTT ACG GAT GGC ATG ACA GTA AGA GAA TTA TGC AGT GCT GCC ATA  
Lys His Leu Thr Asp Gly Met Thr Val Arg Glu Leu Cys Ser Ala Ala Ile  
20 320 330 340 350 360  
\* \* \* \* \*  
ACC ATG AGT GAT AAC ACT GCG GCC AAC TTA CTT CTG ACA ACG ATC GGA GGA  
Thr Met Ser Asp Asn Thr Ala Ala Asn Leu Leu Leu Thr Thr Ile Gly Gly  
370 380 390 400 410  
\* \* \* \* \*  
25 CCG AAG GAG CTA ACC GCT TTT TTG CAC AAC ATG GGG GAT CAT GTA ACT CGC  
Pro Lys Glu Leu Thr Ala Phe Leu His Asn Met Gly Asp His Val Thr Arg  
420 430 440 450 460  
\* \* \* \* \*  
CTT GAT CAT TGG GAA CCG GAG CTG AAT GAA GCC ATA CCA AAC GAC GAG CGT  
30 Leu Asp His Trp Glu Pro Glu Leu Asn Glu Ala Ile Pro Asn Asp Glu Arg  
470 480 490 500 510  
\* \* \* \* \*  
GAC ACC ACG ATG CCT GTA GCA ATG GCA ACA ACG TTG CGC AAA CTA TTA ACT  
Asp Thr Thr Met Pro Val Ala Met Ala Thr Thr Leu Arg Lys Leu Leu Thr  
35 520 530 540 550 560  
\* \* \* \* \*  
GGC GAA CTA CTT ACT CTA GCT TCC CGG CAA CAA TTA ATA GAC TGG ATG GAG  
Gly Glu Leu Leu Thr Leu Ala Ser Arg Gln Gln Leu Ile Asp Trp Met Glu  
40 570 580 590 600 610  
\* \* \* \* \*  
GCG GAT AAA GTT GCA GGA CCA CTT CTG CGC TCG GCC CTT CCG GCT GGC TGG  
Ala Asp Lys Val Ala Gly Pro Leu Leu Arg Ser Ala Leu Pro Ala Gly Trp  
620 630 640 650 660  
\* \* \* \* \*  
45 TTT ATT GCT GAT AAA TCT GGA GCC GGT GAG CGT GGG TCT CGC GGT ATC ATT  
Phe Ile Ala Asp Lys Ser Gly Ala Gly Glu Arg Gly Ser Arg Gly Ile Ile  
670 680 690 700 710 720  
\* \* \* \* \*  
GCA GCA CTG GGG CCA GAT GGT AAG CCC TCC CGT ATC GTA GTT ATC TAC ACG  
50 Ala Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg Ile Val Val Ile Tyr Thr  
730 740 750 760 770  
\* \* \* \* \*  
ACG GGG AGT CAG GCA ACT ATG GAT GAA CGA AAT AGA CAG ATC GCT GAG ATA

Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn Arg Gln Ile Ala Glu Ile  
780 790  
\* \* \* \* \*  
GGT GCC TCA CTG ATT AAG CAT TGG  
5 Gly Ala Ser Leu Ile Lys His Trp

10 SEQ.ID NO. 4: range 1 to 792

15 \* \* \* \* \*  
ATG GAC CCA GAA ACG CTG GTG AAA GTA AAA GAT GCT GAA GAT CAG TTG GGT  
Met Asp Pro Glu Thr Leu Val Lys Val Lys Asp Ala Glu Asp Gln Leu Gly  
60 70 80 90 100  
\* \* \* \* \*  
20 GCA CGA GTG GGT TAC ATC GAA CTG GAT CTC AAC AGC GGT AAG ATC CTT GAG  
Ala Arg Val Gly Tyr Ile Glu Leu Asp Leu Asn Ser Gly Lys Ile Leu Glu  
110 120 130 140 150  
\* \* \* \* \*  
25 AGT TTT CGC CCC GAA GAA CGT TTT CCA ATG ATG AGC ACT TTT AAA GTT CTG  
Ser Phe Arg Pro Glu Glu Arg Phe Pro Met Met Ser Thr Phe Lys Val Leu  
160 170 180 190 200  
\* \* \* \* \*  
30 CTA TGT GGC GCG GTA TTA TCC CGT ATT GAC GCC GGG CAA GAG CAA CTC GGT  
Leu Cys Gly Ala Val Leu Ser Arg Ile Asp Ala Gly Gln Glu Gln Leu Gly  
210 220 230 240 250  
\* \* \* \* \*  
35 CGC CGC ATA CAC TAT TCT CAG AAT GAC TTG GTT GAG TAC TCA CCA GTC ACA  
Arg Arg Ile His Tyr Ser Gln Asn Asp Leu Val Glu Tyr Ser Pro Val Thr  
260 270 280 290 300  
\* \* \* \* \*  
40 GAA AAG CAT CTT ACG GAT GGC ATG ACA GTA AGA GAA TTA TGC AGT GCT GCC  
Glu Lys His Leu Thr Asp Gly Met Thr Val Arg Glu Leu Cys Ser Ala Ala  
310 320 330 340 350  
\* \* \* \* \*  
45 ATA ACC ATG AGT GAT AAC ACT GCG GCC AAC TTA CTT CTG ACA ACG ATC GGA  
Ile Thr Met Ser Asp Asn Thr Ala Ala Asn Leu Leu Leu Thr Thr Ile Gly  
360 370 380 390 400  
\* \* \* \* \*  
50 GGA CCG AAG GAG CTA ACC GCT TTT TTG CAC AAC ATG GGG GAT CAT GTA ACT  
Gly Pro Lys Glu Leu Thr Ala Phe Leu His Asn Met Gly Asp His Val Thr  
410 420 430 440 450  
\* \* \* \* \*  
CGC CTT GAT CAT TGG GAA CCG GAG CTG AAT GAA GCC ATA CCA AAC GAC GAG  
Arg Leu Asp His Trp Glu Pro Glu Leu Asn Glu Ala Ile Pro Asn Asp Glu  
460 470 480 490 500 510  
\* \* \* \* \*  
CGT GAC ACC ACG ATG CCT GTA GCA ATG GCA ACA ACG TTG CGC AAA CTA TTA  
Arg Asp Thr Thr Met Pro Val Ala Met Ala Thr Thr Leu Arg Lys Leu Leu

```

      520      530      540      550      560
      *      *      *      *      *      *
ACT GGC GAA CTA CTT ACT CTA GCT TCC CGG CAA CAA TTA ATA GAC TGG ATG
Thr Gly Glu Leu Leu Thr Leu Ala Ser Arg Gln Gln Leu Ile Asp Trp Met
5      570      580      590      600      610
      *      *      *      *      *      *
GAG GCG GAT AAA GTT GCA GGA CCA CTT CTG CGC TCG GCC CTT CCG GCT GGC
Glu Ala Asp Lys Val Ala Gly Pro Leu Leu Arg Ser Ala Leu Pro Ala Gly
      620      630      640      650      660
10     *      *      *      *      *      *
TGG TTT ATT GCT GAT AAA TCT GGA GCC GGT GAG CGT GGG TCT CGC GGT ATC
Trp Phe Ile Ala Asp Lys Ser Gly Ala Gly Glu Arg Gly Ser Arg Gly Ile
      670      680      690      700      710
      *      *      *      *      *      *
15     ATT GCA GCA CTG GGG CCA GAT GGT AAG CCC TCC CGT ATC GTA GTT ATC TAC
Ile Ala Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg Ile Val Val Ile Tyr
      720      730      740      750      760
      *      *      *      *      *      *
ACG ACG GGG AGT CAG GCA ACT ATG GAT GAA CGA AAT AGA CAG ATC GCT GAG
20     Thr Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn Arg Gln Ile Ala Glu
      770      780      790
      *      *      *      *      *
ATA GGT GCC TCA CTG ATT AAG CAT TGG
Ile Gly Ala Ser Leu Ile Lys His Trp
25

```

30 SEQ.ID NO. 5: range 1 to 786

```

      10      20      30      40      50
      *      *      *      *      *      *
35     ATG AAA GAT GAT TTT GCA AAA CTT GAG GAA CAA TTT GAT GCA AAA CTC GGG
Met Lys Asp Asp Phe Ala Lys Leu Glu Glu Gln Phe Asp Ala Lys Leu Gly
      60      70      80      90      100
      *      *      *      *      *      *
40     ATC TTT GCA TTG GAT ACA GGT ACA AAC CGG ACG GTA GCG TAT CGG CCG GAT
Ile Phe Ala Leu Asp Thr Gly Thr Asn Arg Thr Val Ala Tyr Arg Pro Asp
      110      120      130      140      150
      *      *      *      *      *      *
45     GAG CGT TTT GCT TTT GCT TCG ACG ATT AAG GCT TTA ACT GTA GGC GTG CTT
Glu Arg Phe Ala Phe Ala Ser Thr Ile Lys Ala Leu Thr Val Gly Val Leu
      160      170      180      190      200
      *      *      *      *      *      *
50     TTG CAA CAG AAA TCA ATA GAA GAT CTG AAC CAG AGA ATA ACA TAT ACA CGT
Leu Gln Gln Lys Ser Ile Glu Asp Leu Asn Gln Arg Ile Thr Tyr Thr Arg
      210      220      230      240      250
      *      *      *      *      *      *
GAT GAT CTT GTA AAC TAC AAC CCG ATT ACG GAA AAG CAC GTT GAT ACG GGA
Asp Asp Leu Val Asn Tyr Asn Pro Ile Thr Glu Lys His Val Asp Thr Gly
      260      270      280      290      300

```

097401350

```

      *      *      *      *      *      *      *      *      *
ATG ACG CTC AAA GAG CTT GCG GAT GCT TCG CTT CGA TAT AGT GAC AAT GCG
Met Thr Leu Lys Glu Leu Ala Asp Ala Ser Leu Arg Tyr Ser Asp Asn Ala
      310      320      330      340      350
5      *      *      *      *      *      *      *      *      *
GCA CAG AAT CTC ATT CTT AAA CAA ATT GGC GGA CCT GAA AGT TTG AAA AAG
Ala Gln Asn Leu Ile Leu Lys Gln Ile Gly Gly Pro Glu Ser Leu Lys Lys
      360      370      380      390      400
      *      *      *      *      *      *      *      *      *
10     GAA CTG AGG AAG ATT GGT GAT GAG GTT ACA AAT CCC GAA CGA TTC GAA CCA
      Glu Leu Arg Lys Ile Gly Asp Glu Val Thr Asn Pro Glu Arg Phe Glu Pro
      410      420      430      440      450
      *      *      *      *      *      *      *      *      *
15     GAG TTA AAT GAA GTG AAT CCG GGT GAA ACT CAG GAT ACC AGT ACA GCA AGA
      Glu Leu Asn Glu Val Asn Pro Gly Glu Thr Gln Asp Thr Ser Thr Ala Arg
      460      470      480      490      500      510
      *      *      *      *      *      *      *      *      *
20     GCA CTT GTC ACA AGC CTT CGA GCC TTT GCT CTT GAA GAT AAA CTT CCA AGT
      Ala Leu Val Thr Ser Leu Arg Ala Phe Ala Leu Glu Asp Lys Leu Pro Ser
      520      530      540      550      560
      *      *      *      *      *      *      *      *      *
25     GAA AAA CGC GAG CTT TTA ATC GAT TGG ATG AAA CGA AAT ACC ACT GGA GAC
      Glu Lys Arg Glu Leu Leu Ile Asp Trp Met Lys Arg Asn Thr Thr Gly Asp
      570      580      590      600      610
      *      *      *      *      *      *      *      *      *
30     GCC TTA ATC CGT GCC GGA GCG GCA TCA TAT GGA ACC CGG AAT GAC ATT GCC
      Ala Leu Ile Arg Ala Gly Val Pro Asp Gly Trp Glu Val Ala Asp Lys Thr
      620      630      640      650      660
      *      *      *      *      *      *      *      *      *
35     ATC ATT TGG CCG CCA AAA GGA GAT CCT GTC GGT GTG CCG GAC GGT TGG GAA
      Gly Ala Ala Ser Tyr Lys Gly Asp Pro Val Gly Thr Arg Asn Asp Ile Ala
      670      680      690      700      710
      *      *      *      *      *      *      *      *      *
      GTG GCT GAT AAA ACT GTT CTT GCA GTA TTA TCC AGC AGG GAT AAA AAG GAC
      Ile Ile Trp Pro Pro Val Leu Ala Val Leu Ser Ser Arg Asp Lys Lys Asp
      720      730      740      750      760
      *      *      *      *      *      *      *      *      *
40     GCC AAG TAT GAT GAT AAA CTT ATT GCA GAG GCA ACA AAG GTG GTA ATG AAA
      Ala Lys Tyr Asp Asp Lys Leu Ile Ala Glu Ala Thr Lys Val Val Met Lys
      770      780
      *      *      *      *
      GCC TTA AAC ATG AAC GGC AAA
      Ala Leu Asn Met Asn Gly Lys

```

45